(NASA-STD-8070.5A) NASA STANDARD: TREND ANALYSIS TECHNIQUES (NASA) 91 p CSCL 12A

N91-22815

Unclas

H1/65 0008473

PREFACE

This Standard, first published in October 1988, was prepared as an aid to scientists, engineers and managers who are responsible for identifying and interpreting trends in NASA programs.

In this first revision to the original Standard, material covering common trending errors and tests of significance has been expanded. Sections 4.4.9 on Normalization of Trend Data, 4.6 on Potential Problems in Determining and Analyzing Trends, as well as paragraphs on R^2 values and significance of fit have been included in this revision. It is hoped that this additional material will improve the understanding and use of trend techniques throughout NASA.

Data and examples in this document have been developed to aid in illustrating analytic techniques and should not be construed to represent actual data or trends unless specifically noted.

Questions should be addressed to NASA Headquarters, Systems Assessment and Trend Analysis Division (Code QT), Office of the Associate Administrator for Safety and Mission Quality, Washington, DC 20546.

Michael A. Greenfield

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

Figure **Page**

LIST OF TABLES

Table **Page**

Chapter 1

PURPOSE AND SCOPE

This Standard presents descriptive and analytical techniques for NASA trend analysis applications. Trend analysis is applicable in all organizational elements of NASA connected with, or supporting, developmental/operational programs. Use of this Standard is not mandatory; however, it should be consulted for any data analysis activity requiring the identification or interpretation of trends.

Trend Analysis is neither a precise term nor a circumscribed methodology, but rather connotes, generally, quantitative analysis of time-series data. For NASA activities, the appropriate and applicable techniques include descriptive and graphical statistics, and the fitting or modeling of data by linear, quadratic, and exponential models. Usually, but not always, the data is time-series in nature. Concepts such as autocorrelation and techniques such as Box-Jenkins time-series analysis would only rarely apply and are not included in this Standard.

The document presents the basic ideas needed for qualitative and quantitative assessment of trends, together with relevant examples. A list of references provides additional sources of information.

Chapter 2

BACKGROUND

2.1 REPORTS BY ROGERS COMMISSION AND SLAY COMMISSION

The Rogers Commission was appointed by the President to investigate the Challenger Space Shuttle accident. The Commission's mandate was to determine the probable cause(s) of the accident; and, based on its investigative findings, to provide recommendations for improvements in space program safety. The committee's consensus was that sealing failure of O-rings in the Shuttle's right Solid Rocket Booster was the specific failure leading to the accident. Investigation of the circumstances surrounding the accident revealed that data on O-ring anomalies in previous shuttle flights showed a hazardous trend beginning in January 1984. However, no quantitative trend analysis was conducted prior to mission 51-L to reveal and substantiate this trend. Thus, the developing trend in O-ring erosion/blowby went unnoticed.

In response to the Committee recommendations, the Office of Safety, Reliability, Maintainability and Quality Assurance was established. (As of October 1990, the name was changed to the Office of Safety and Mission Quality.) The Office functions in a policy determination and oversight capacity for all NASA SRM&QA activities, and has responsibility for ensuring that trend analysis techniques are applied, where appropriate, in NASA programs.

In September of 1986, the committee on Shuttle Criticality Review and Hazard Analysis Audit, known as the Slay Commission, was established in response to a recommendation by the Rogers Commission. The Committee's purpose was to evaluate NASA efforts in reviewing safety-critical shuttle components and identifying components needing improvement. The Committee's report was more broadly focused, evaluating the entirety of STS risk assessment and risk management procedures.

The Slay Commission recommended that NASA employ more quantitative methods and measures in assessing STS risks, and in supporting NASA management decisions regarding the STS. Trend analysis is one quantitative technique to identify potentially hazardous conditions based on past empirical data.

2.2 RELEVANT NASA GUIDANCE

NASA Management Instructions 8070.3 and 8070.4 pertain to specific trend analysis requirements and risk management policies, respectively. NASA Handbook 8070.5 is a companion document to NMI 8070.3 and further defines the responsibilities, classifications, and overall guidelines for the conduct of trend analysis for manned spacecraft and associated payloads.

2.3 MOTIVATION FOR TREND ANALYSIS - BASIC OBJECTIVE AND BENEFITS

In assessing system (or component) reliability and safety, the most complete *quantitative* knowledge would be in a reliability function (or equivalently, a risk function). The reliability of a system (or component) is a real-valued function with range in the interval [0,1]; it is a *probability function*, giving the probability of adequate performance under specified conditions up to a time, t. Usually this reliability function, denoted R, is a function of a time variable and certain parameters. For example, a wear-out time-to-failure model widely used is the Weibull model $R = \exp(-at^b)$, where *a*

and b are parameters to be estimated. More generally, sometimes the parameters of the reliability function are themselves functions of other variables, such as pressure and temperature.

Determining the reliability distribution and estimating its parameters may be very difficult, costly, and perhaps intractable. **Trend analysis** is an alternative or companion approach. Specifically, we know that the values of certain variables will directly impact on a component or system's reliability, even though the exact quantitative relationship or risk has not been determined. Those measurable variables (parameters) that directly affect system or component reliability are sampled *over time*. The variable values are examined to see if there is a pattern of deviation over time (i.e., a **trend**) from acceptable performance limits. In this manner, one may be able to predict future parameter values, or at least estimate the long-term range of values of these influential variables. In turn, if these parameters are trending towards hazardous or unacceptable levels, the potential problem could be identified *prior* to the occurrence of high-risk situations.

The preceding trending approach is termed **performance trending**. In other applications, the focus may be less on detecting a deviation from an acceptable limit than on detecting an overall upward or downward movement of a set of observed values over time. This general trending could apply to programmatic or reliability concerns: for example, significant problem reports; workmanship defects; cannibalizations; relative risks and failure rates. Trending applied to management indicators is termed **programmatic trending**. When applied to frequency of occurrence, it is termed **problem trending. Supportability trending** focuses on specific subject matter, namely logistics information. In all of these domains, a trend analysis can reveal a movement towards unacceptable, undesirable or dangerous reliability, safety, or quality assurance levels. Also, if the particular trending model has a significant quantitative fit (e.g., linear, quadratic, exponential), future predictions can be made.

From a management or reliability point of view, there are cases where a distinct trend should exist. For example, there should be an increasing trend in reliability with successive design changes and a decreasing trend in open significant problem reports as time-to-launch decreases.

Thus, measurable variables will impact, directly or indirectly, system reliability, safety, and quality assurance. Even though the exact or approximate quantitative relationship to reliability may not be known, following the trending of these variables may identify *potentially* significant reliability/safety problems. This *proactive* posture is both the objective and tangible benefit of the trend analysis approach.

Chapter 3

TREND ANALYSIS - OVERVIEW

3.1 FOUR CLASSES OF TREND TYPES

NASA Handbook 8070.5 provides the definitions for the four types of Trend Analysis: Performance, Problem, Programmatic, and Supportability. The following summarizes these definitions:

a. ***Performance Trending***. This trend analysis is used to measure a condition(s) that is changing over time in a manner that eventually will cause the part/system to fail. This technique can be considered a simple reliability model. A simplified example of a performance trend would be the decline in electrical output of a deployed solar panel over time.

b. ***Problem Trending***. This trend analysis tracks and categorizes problems over time. The problems may be for an entire system, subsystem, or any other appropriate level of aggregation. This technique allows managers to focus on the problems that are occurring with the most frequency.

c. ***Programmatic Trending***. This trend analysis quantitatively monitors program-related indicators such as overtime, critical schedule elements, and manpower resource availability, which relate to or reflect a potential impact on safety or mission/operational success.

d. ***Supportability Trending***. This trend analysis quantitatively examines over time logistic support elements. The trending areas for supportability would include such issues as servicing, repair, spares, overhaul and refurbishment, etc.

3.2 BASIC DEFINITIONS OF STATISTICAL TERMS

Abscissa	X coordinates in a rectangular coordinate system.
Alpha Error (α)	The probability of rejecting a true population parameter or average. Also called Type I error or Producer's Risk (rejected a good lot).
Beta Error (β)	The probability of accepting a false or incorrect population parameter or average. Also called a Type II error or Consumer's Risk (accepted a bad lot).
Confidence Interval	The interval computed around an estimated parameter, which expresses the probability of including the true population value within its bounds.
Correlation	The degree of relationship between two variables.
Extrapolation	A prediction of a Y value using an X value outside of the X range from which the model was derived.

Frequency Distribution	A tabular or graphical arrangement of data by classes, along with the corresponding class frequencies.
Independence	If the occurrence or nonoccurrence of one event does not affect the probability of occurrence of a second event, then these are independent events.
Intercept	The Y value when $X = 0$ for a line plotted in the X-Y coordinate system.
Interpolation	Estimating a Y value between two known (X,Y) pairs.
Least Squares	In cases where a linear relationship is known to exist, or may be reasonably assumed to exist, the principle of minimizing the square of the residuals is called least squares. Fitting a regression line through data points by minimizing the sum of the squares from the fitted line to the observed points.
Mean	The term used to describe a population or sample average. For a variable X, the mean is usually denoted by $\overline{X}$ or μ.
Normal Distribution	A type of symmetrical or bell-shaped frequency curve characterized by the fact that observations equidistant from the mean have the same frequency.
Ordinate	Y coordinates in a rectangular coordinate system.
Parameters	The term applied to population or sample characteristics such as the mean and standard deviation.
Pareto	The concept that a relatively large percentage (80-90%) of problems will be caused by a relatively small percentage (10-20%) of related factors.
Pareto Diagram	A rank ordering of problem causes by their contribution, usually in decreasing order.
Population	Any group of items. A universe.
Projection	An expansion of sample results to population values.
Range	The difference between the largest and smallest item in a set of data.
Residuals	Also known as *random disturbance*. The *error* in fitting a line/curve through a set of data points.
Risks	The term applied to either or both the alpha or beta error.
Sample Size	The number of items selected from a population that will be used to make inferences about the total population.

Significance Level	Same as alpha error. Computed as 1 - (confidence level).
Slope	The rate of change in Y per unit change in X for a line plotted in the X-Y coordinate system.
Standard Deviation	A measure of variability used in common statistical tests. It is the square root of the variance and is usually denoted by σ.
Universe	Same as population. The total group of items possessing a certain characteristic(s).
Variability	A term expressing the spread of items around a sample average.
Variance	A measure of the spread or variability in a set of data. The average of the sum of squares of individual deviations from the mean and is usually denoted by σ^2.

Chapter 4

ANALYTIC AND DESCRIPTIVE TREND TECHNIQUES; APPLICABILITY; EXAMPLES

4.1 DESCRIPTIVE STATISTICS

Descriptive statistics refers to raw data frequencies and tabulations, and simple measures such as the mean (average), median, and percentiles. While these quantifications are the most fundamental and elementary, they are probably the most important approaches in the NASA efforts for trend analysis. *More sophisticated analyses should be preceded by the generation and examination of basic descriptive statistics. In many cases, a descriptive statistics approach, coupled with a graphical portrayal of the data, will be sufficient for trending purposes.*

4.1.1 Frequencies and Tabulations

The **frequency** of a variable is simply a numerical count of the distinct values (or groups of values) for the variable. A standard application would be the number of problem reports from the PRACA Database on components of a subsystem (e.g., the STS ET) for a given time period; or the number of occurrences of each failure mode for a specific STS hardware component. In many cases, the first analysis step will be a frequency for the highest failing or highest problem-related criticality 1/1R/1S components (usually the top 10-25).

A **2-way tabulation** is an n x k frequency table. Tabulations of this type are useful in investigating and portraying the frequency of occurrence for a pair of classifications or variables. In the trending effort, tabulation for the highest problem-related criticality 1/1R/1S parts for the current STS prelaunch period, and their respective counts for past periods, is useful to see if there is any developing trend. Such a tabulation would be of the form:

Highest Problem-Related CRIT 1/1R Components for Current Prelaunch Period
Subsystem = _____

Component	Criticality	Launch Periods			
		Current	-1	-2	-3

Note: Highest Problem-Related Components are determined for current period and historical frequencies are traced for previous periods to display possible trends.

4.2 GRAPHICAL TECHNIQUES

The purpose for any collection of data is to convert data into information. This information is then used in support of management decisions. Graphical techniques provide simple, yet clear and concise means to transform raw data into information. A representative sample of the most commonly used techniques for quality assurance and reliability are presented in the following paragraphs. These techniques may be easily modified or combined to portray information.

4.2.1 Scatterplots

As emphasized throughout this Standard, one of the first steps in analyzing any set of data is to plot it. The first type of plot considered for trend data is a scatterplot. This plot is simply a bivariate plot with time on the x axis and the corresponding values on the y-axis. Additional information such as the regression line, moving averages or other smoothing procedures, and additional sets of y values, can be overlaid on this plot.

Figure 4.2A is a sample scatterplot that shows the percent of welds failing inspection over a 16-week time period.

Figure 4.2A. Scatterplot (Sample)

4.2.2 Control Charts

Control charts are the basic graphical tool used in quality control. They are an extremely efficient way to present a large amount of information. There are numerous types of control charts, several of which are listed in Chapter 4, Section 4.5.1. Figure 4.2B is a sample control chart that is applicable for the thickness of an item such as an orbiter tile or wall thickness of a heat exchanger.

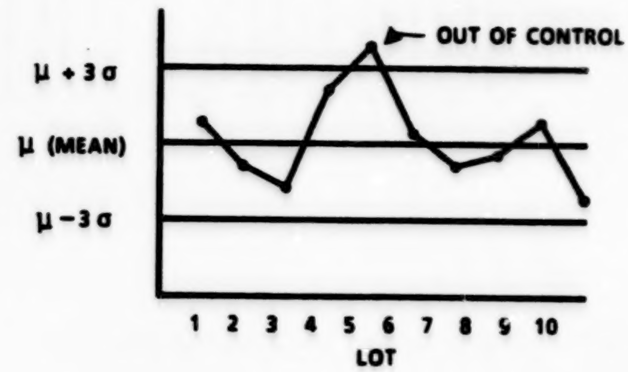


Figure 4.2B. Control Chart (Sample)

4.2.3 Bar Plots

Bar plots are commonly used to portray changes over time, and to compare more than one set of data on the same graph (paired bar graphs). Through the use of desktop computer software packages, three-dimensional bar plots provide further resolution to the graphical data.

Figures 4.2C through 4.2E are examples of bar plots.

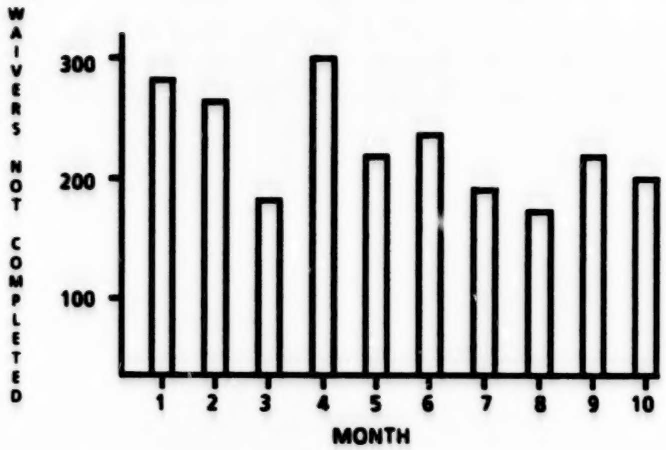


Figure 4.2C. Bar Plot (Sample)

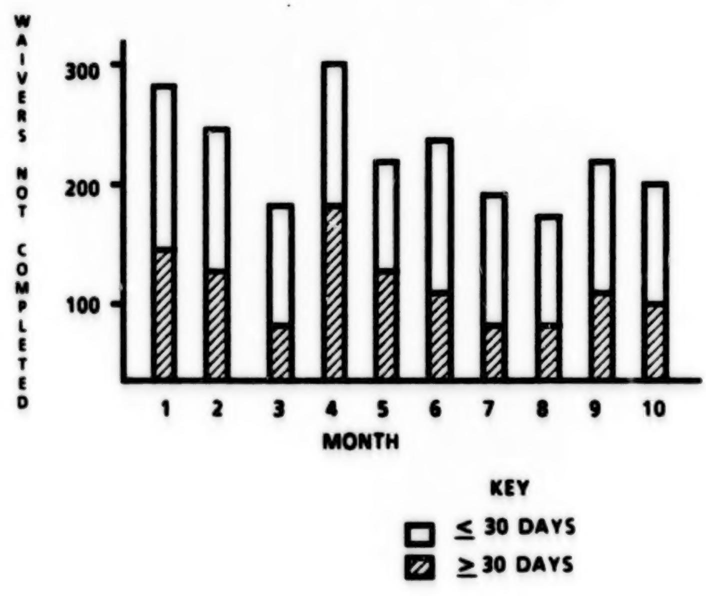


Figure 4.2D. Segmented Bar Graph (Sample)

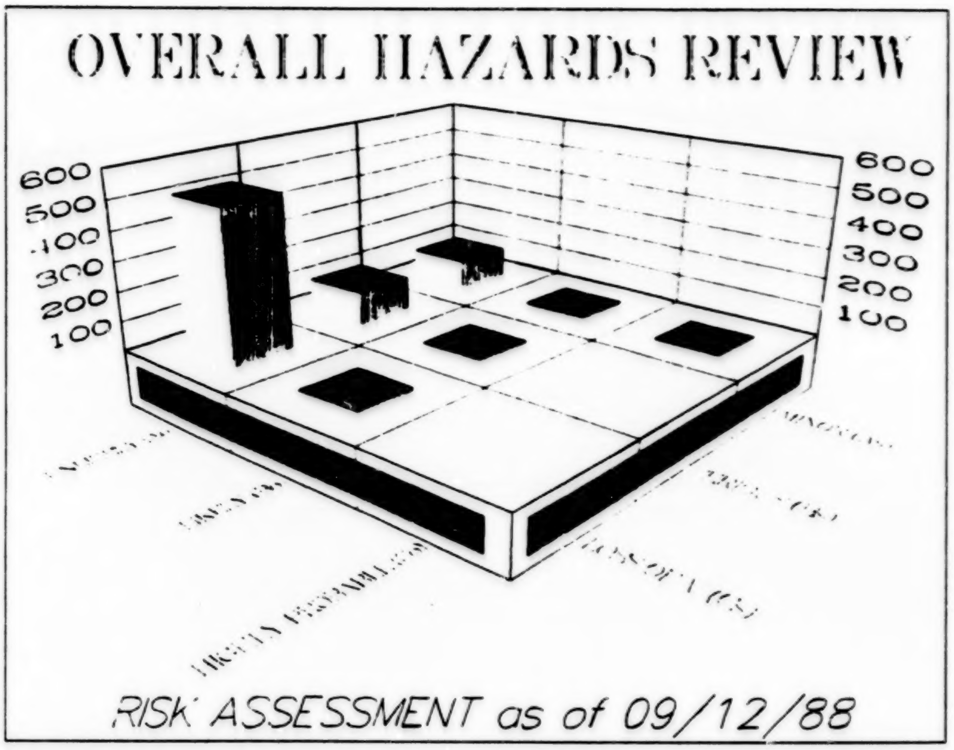
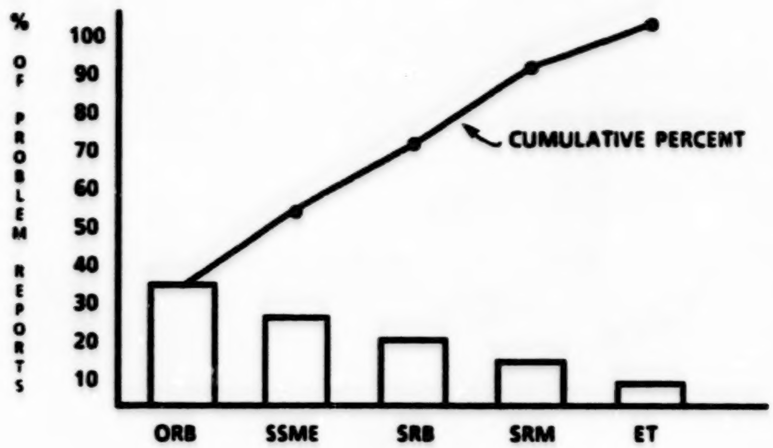


Figure 4.2E. Three-Dimensional Bar Plot (Sample)

4.2.4 Pareto Charts

Pareto charts place emphasis on the type of defect responsible for the most problems. As with all graphical techniques, there are many variations. *In its simplest form, the Pareto Chart is a sorted bar chart.* However, by using segmented bars (see figure 4.2D) and/or time period comparisons (see figure 4.2G), much more information can be conveyed than through a simple bar graph. Note that when using pareto charts for a *before and after* comparison, actual data rather than percent must be plotted to provide an accurate picture of the comparison (i.e., if welds are 50% of 100 problems and you reduce the number of problems to 30, with 15 being welds, a percentage pareto chart will show no improvement for weld errors when, in fact, they have decreased from 50 to 15).



(*NOTE*: BARS ARE OFTEN SEGMENTED)

Figure 4.2F. Typical Pareto Chart

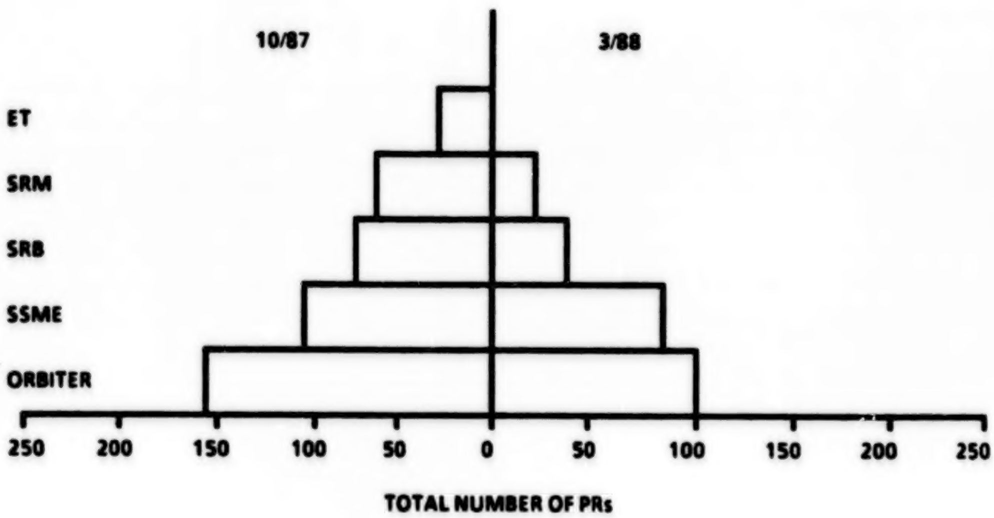


Figure 4.2G. Pareto Pyramid Chart

4.2.5 Histogram (Stem and Leaf)

A **histogram** is a chart that displays frequency of occurrence. Such charts are essential when determining a probability distribution for the data. Histograms are developed by dividing the range of the data into equal intervals, counting the number of data points in each interval, and plotting these counts. A **stem and leaf plot** provides more information than a histogram from the same data. Instead of plotting points, the *actual numerical data values* are used to graphically represent the frequencies. For example, suppose that we want to determine if the distribution of a particular process is approximately normal. Figure 4.2H shows shaft diameters ranging from 1.016" to 1.070".

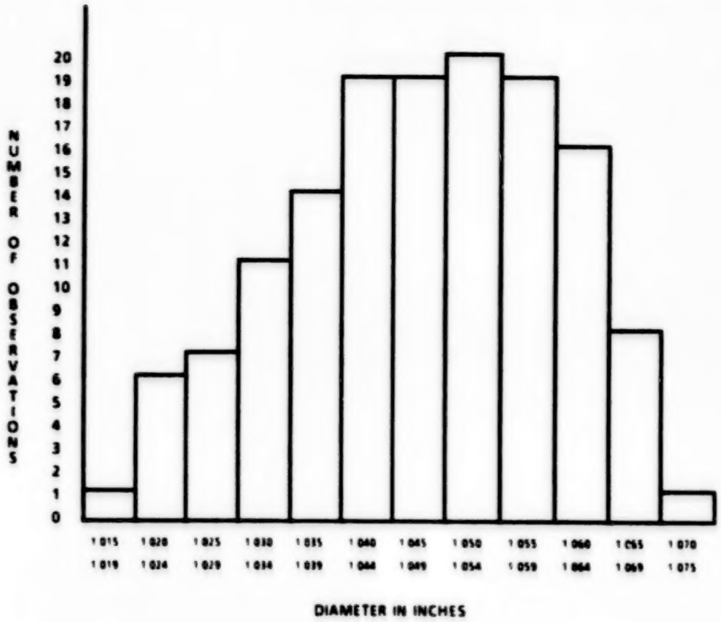


Figure 4.2H. Histogram (Sample)

The same data shown is figure 4.2.H also can be displayed as a stem and leaf plot as shown below:

1.01	6
1.02	011234
1.02	6688899
1.03	01122333444
1.03	55667777788899
1.04	00000111122233334444
1.04	55566777778889999999
1.05	011111122333333333344
1.05	5555566667778888999
1.06	0000011122233334
1.06	56668899
1.07	0

The first line represents 1.016, the second line 1.020, 1.021, 1.021, 1.022, 1.023, and 1.024.

4.2.6 Smoothing

Smoothing is a technique used to adjust time-series data for characteristic movements or variations. These techniques are usually classified into four main types: long-term or secular, cyclical, seasonal, and irregular or random. The methods that are used to smooth data include moving averages, moving medians, moving midpoints, and exponential smoothing. The exponential and moving average techniques are the most applicable and are developed in the next section.

Figure 4.21 shows how smoothed data allows one to visually discern long-term trends and data patterns that are obscured by fluctuations in the raw data.

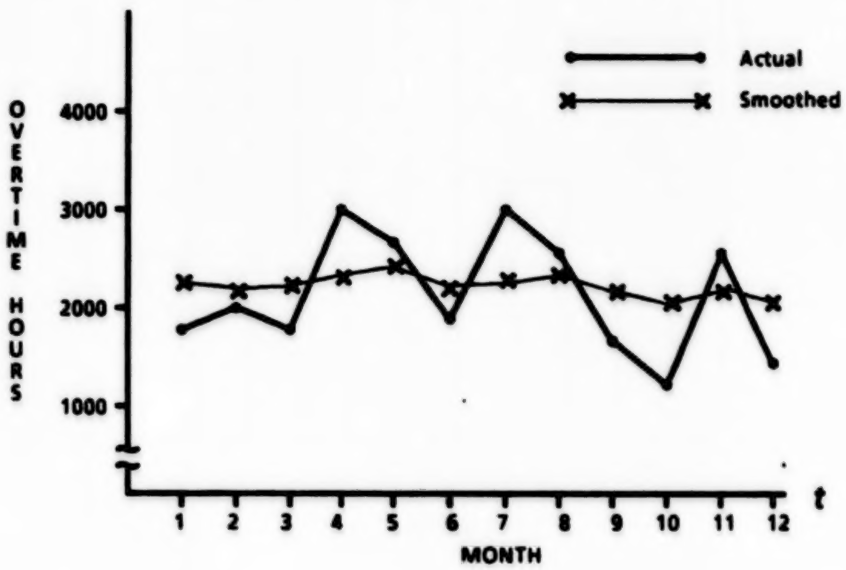


Figure 4.21. Smoothed Data (Sample)

4.2.7 Box Plots

Box plots are an effective means of summarizing information from a large amount of data. When done with time on the x-axis, they are useful in visually presenting trend data. The construction of a simple box plot without the x-axis is shown in figure 4.2J.

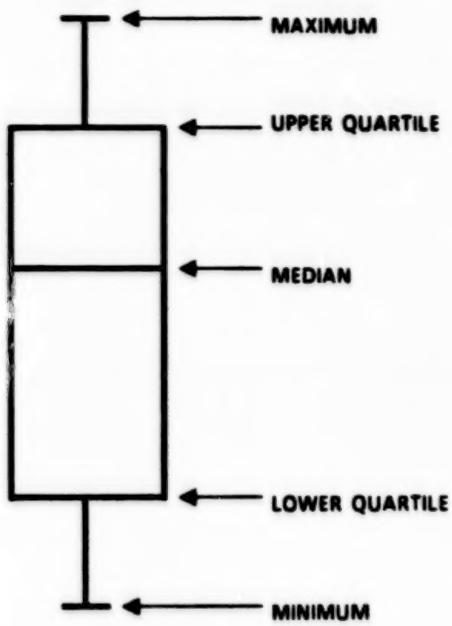


Figure 4.2J. Box Plot (Sample)

More sophisticated box plots include: notching the box around the median in proportion to the median's variability; varying the width to indicate sample size; and plotting outliers more than 1.5 times the interquartile distances from either quartile as asterisks.

4.2.8 Location Plots

Location plots graphically represent the location and the type of error/defect occurring in a process. These charts are maintained over time; therefore, certain trends usually can be observed as they develop. For example, a circuit card may exhibit the pattern of solder defects, as shown in figure 4.2K.

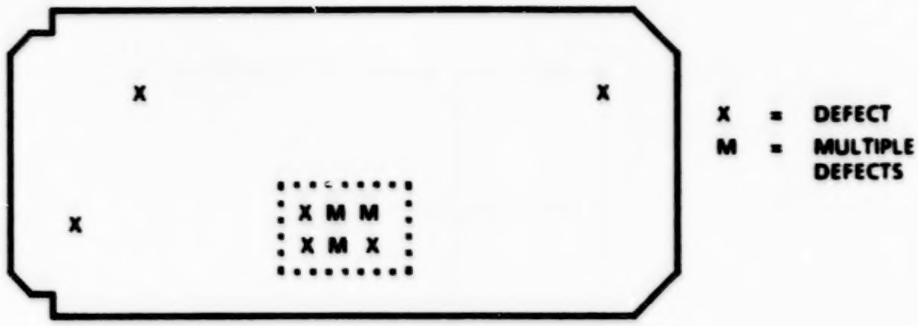


Figure 4.2K. Circuit Card Solder Defects

The area surrounded by the dotted line is a candidate for a thorough investigation. Redesign may be necessary to enable this soldering to be achieved with higher quality. The chart also may be drawn using more descriptive and multiple indicators for defects, such as defect type, machine operator shift, and day of week.

4.2.9 Examples

4.2.9.1 Example 1: Goddard Satellite Missions - 1986 Anomalies. There have been 49 Goddard satellite missions since 1970. The following Pareto charts aid in classifying the anomalies for satellites in orbit in 1986, and hence, the possible general failure modes. These satellites are:

ERBS	NOAA-6
GDES-4	NOAA-9
GDES-5	NOAA-10
GDES-6	SMM
LANDSAT-5	TDRS-1
NIMBUS-7	

The Pareto chart approach is used to globally ascertain the problem areas and rank problems by frequency. Figures 4.2L through 4.2O identify satellite anomalies by: satellite subsystem; criticality of the failure; type of failure mode; and impact of the failure on satellite functioning.

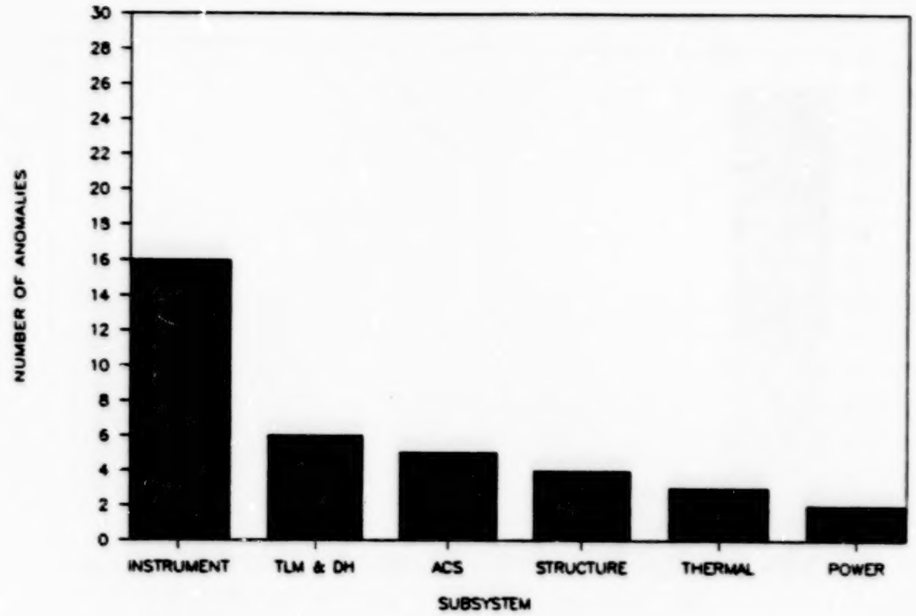


Figure 4.2L. 1986 Anomalies by Subsystem

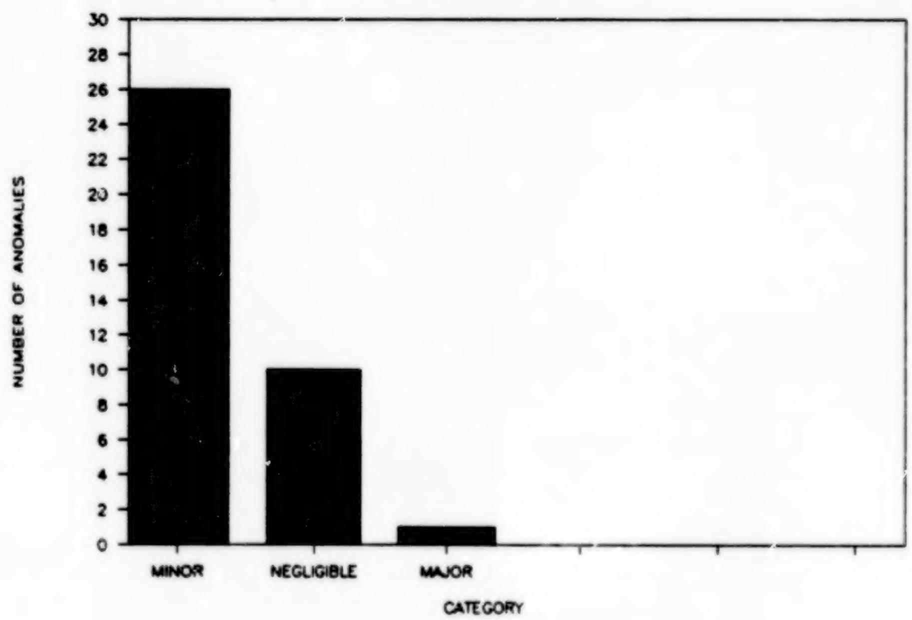


Figure 4.2M. 1986 Anomalies By Criticality (Mission Effect)

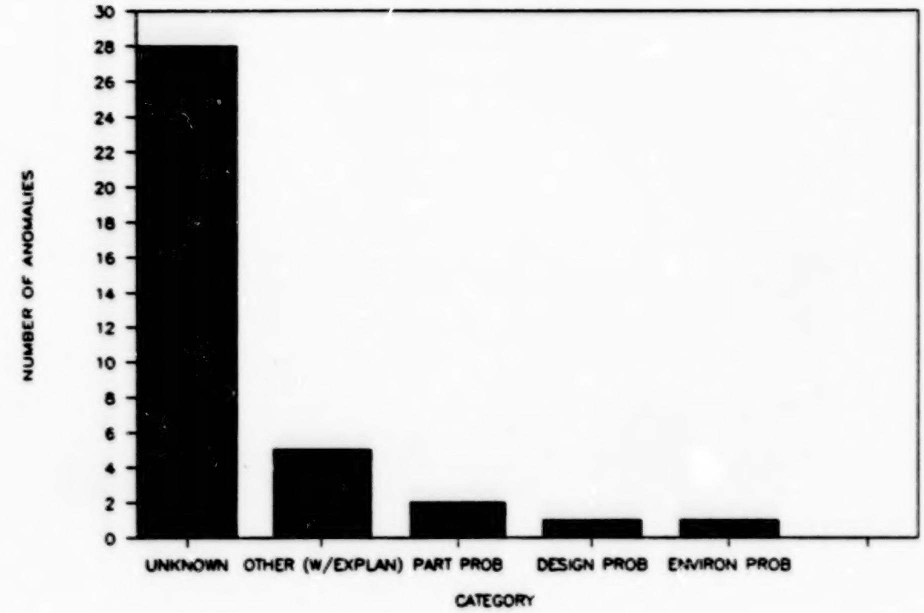


Figure 4.2N. 1986 Anomalies By Failure Category

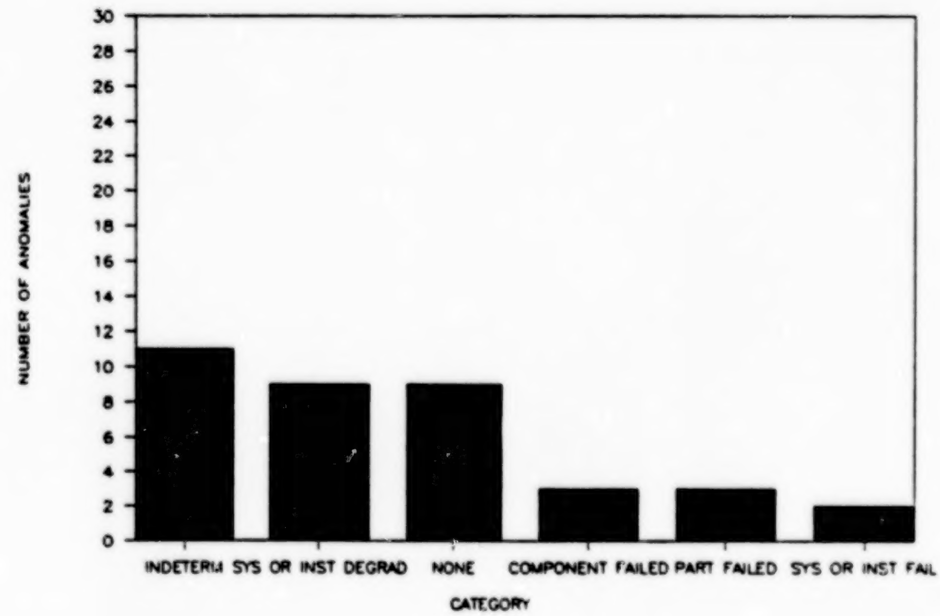
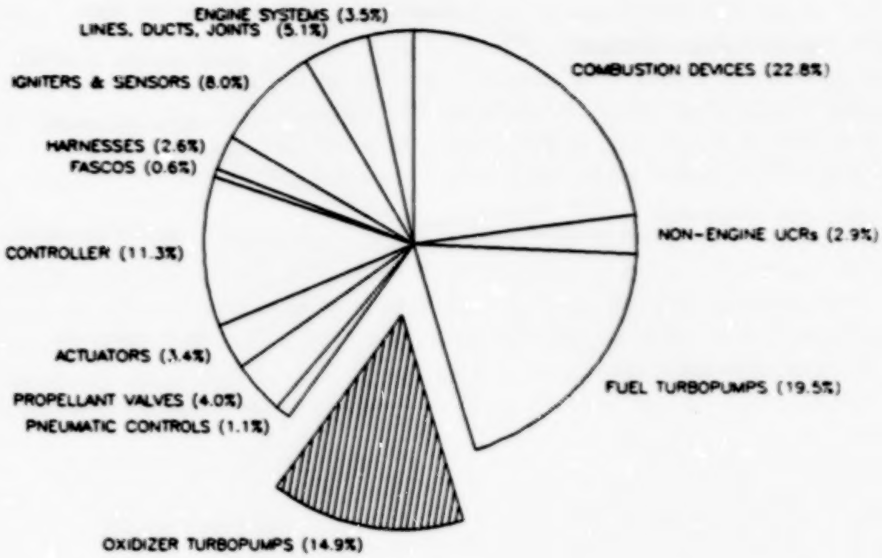


Figure 4.2O. 1986 Anomalies By Impact

4.2.9.2 Example 2: Distribution of SSME UCRs/HPOT; FMEA Attributable UCRs. The following pie chart shows the distribution of Unsatisfactory Condition Reports (UCRs) written against shuttle main engines from January 1981 through December 1987. UCRs report both failures and other unsatisfactory conditions or potential problems; the reports follow SSME components from manufacturing acceptance tests throughout the component life-cycle. During the January 1981 through December 1987 period, 6,002 UCRs were written against SSMEs, which encompassed 25 shuttle launches, 764 engine tests, and 219,311 seconds of engine operation.

Restricting attention to the High Pressure Oxidizer Turbopump (HPOT), 730 UCRs were written, 217 of which could be directly related to FMEA/CIL failure modes. (Other UCRs referenced design/development, fabrication/quality, non-failure condition, etc.) The table summarizes the type of failure (referenced by FMEA code) on a quarterly basis. Turbine blade cracking/damage, bearing problems, and turbine disk failures accounted for 38.2%, 21.2%, and 17.5% of the FMEA applicable UCRs, respectively. No increasing or decreasing trends are evident with respect to these kinds of HPOT problems.



	UCRs
Combustion Devices	1370
Fuel Turbopumps	1173
Oxidizer Turbopumps	894
Pneumatic Controls	66
Propellant Valves	243
Actuators	205
Controller	680
Harnesses	159
Igniters and Sensors	462
Lines, Ducts, Joints and Orifices	305
Fascos	35
Engine Systems	213
Non-Engine UCRs	177

Figure 4.2P. All SSME Components Proportion of UCRs (January 1981 - December 1987)

Remark: In general, bar charts are preferable to pie charts. When using pie charts, the relative percentages (or absolute frequencies) of each segment should be given.

Table 4.2-1

SUMMARY OF FMEA MODES AND OTHER UCRs
HPOT (B400) - DISTRIBUTION OF UCRs

	1981				1982				1983				1984				1985				1986				1987				
FMEA MODES UCRs/CRIT*	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	TOTAL
02B Nozzle Vane Damage/1R										1				1	1			1				1	1						6
03A Turbine Blade Cracks/1				1	2									1			1		2	2	1	1		1	2				14
03C Honeycomb Retainer Failure/1																		1											1
03D Turbine Blade Impact Damage/1																			1			1	1						3
03F Loss of Damper Function/1															2	3	1	2	2			1							11
04C Blade Shroud Chipping/1R								2	2	2	2	3	4	3	3	1	4	5	6	1	1	3	1	2	4	3	2		54
05A Interstage Seal Damage/1R													1					1											2
06B Bending, Cracking of Turbine Vanes/1R											1																		1
07F Turbine Disk Failure/1					1	4	2	1	2	1	1	5	2	2	3	2	1		6	2				1	1	1			38
08A Fracture, Distortion of Inlet Vane/1R											2				1				1			2							6
09B Impeller Blade Damage/1R																				3		2		1					6
09C Balance Cavity Leakage/1R																						1							1
13A Bearing Failure/1		1	1	4	6	8	2	3	2	1	3			1		2	2			3	1		1		1	3			45
13E Loss of Bearing Nut Preload/1																					1								1
16A Excessive Primary Secondary Turbine Seal Leakage/1R																											1		1
17A Labyrinth Seal Leakage/3				1							1																		2
20A Fracture/Blockage of Coolant Circuits/1										1	1																		2
20B Coolant Passage Cracks/1																				1									1
21A Weld, Parent Metal, or Drain Line Failure/1									1	3		1		2															7
23A Turbine Piece Part Structural Failure/1			1		1										2	3	1			1		1			1				11
25A Leakage Past the Inboard OPB/HPOTP Seal/3																	1					1		1	1				4

* "WORST CASE" CRITICALITY

4.3 SMOOTHING TIME-SERIES DATA TO PREDICT NEXT-EVENT FREQUENCY

Frequencies or number-of-events from past time periods are often used as predictors of the frequency (or value) for the next time period. For example, the number of cannibalizations for previous shuttle flights should be reliable predictors for cannibalizations in the next flight, assuming conditions remain the same.

Two methods are given for forecasting the next-event value based on past values. These methods are exponential smoothing and moving averages. With exponential smoothing, past event values are weighted by varying powers of a constant α, $0<\alpha<1$; more recent past events are given more weight than events occurring further back in time. The moving averages method has two purposes: (1) as a predictor of the next-event value, a simple average of the previous k events is computed and used as the forecast; and (2) time-series data can be *smoothed* so that an overall upward or downward trend becomes more apparent. Both methods are described in more detail in the following paragraphs.

4.3.1 Exponential Smoothing

Exponential smoothing is a method used to *average* past quantities to predict the value or quantity for the next (future) time period. The prediction or forecasted value for the next time period is simply a weighted average of the *actual* observed value for the current time period and the previously *forecasted* value for the current time period. The method is defined as follows:

Let X_n denote the actual value for the time period, n. By convention, the subscript n will denote the current time period, n-1 will denote the previous time period, and n+1 will denote the next time period. Then,

$$F_{n+1} = \alpha X_n + (1-\alpha) F_n \tag{1}$$

where $0<\alpha<1$, F_n is the value forecasted for period n, and F_{n+1} is the projected forecast for the next time period.

Since F_n is itself a weighted sum of the previous period actual value, and the previous period forecasted value, it follows that an alternative expression for the projected forecast, F_{n+1} is:

$$F_{n+1} = \alpha X_n + \alpha (1-\alpha) X_{n-1} + \alpha (1-\alpha)^2 X_{n-2} + \ldots \tag{2}$$

Thus, the next-period forecast gives a weight to each past observed value, with decreasing weights for observations in the more distant past. Observations in more recent time periods are weighted more heavily (i.e., have a greater contribution to generating a forecast).

Remarks: (1) One problem in using exponential smoothing is to determine a suitable weighting value, α. Note first that a small value for α implies a slow response to change, and that a large value of α (i.e., a value close to 1) gives extreme weight to the most recent event(s). One way to choose a value for α is a *retrospective simulation*: consider all time periods before n; compute a forecasted value for the time period n for different values of α; then choose that value, α_0, which gives a forecast (F_n) that is closest to the observed actual value X_n; use this value, α_0, as the weighting coefficient for forecasting the next time period value, F_{n+1}.

(2) Since the forecasted value will lie on the line segment between the actual value, X_n, and the previously forecasted value, F_n, the next forecast will lag behind any continuing trend. Using the trend slope, the previous forecast can be modified to offset the lag: letting b denote the slope of the trend line, then

$$F_{n+1} = \alpha X_n + (1-\alpha)\ [F_n + b] \tag{3}$$

gives an exponential smoothing adjusted for trend.

(3) There are several different conventions or variations in the calculation of the trend slope, b. Exponential smoothing adjusted for trend is a specific case of the ARIMA (autoregressive integrated moving average) technique. Standard texts on time-series analysis will provide more details on this method.

4.3.2 Moving Averages Method

Given past time-period measurements, the moving-averages method can be used to forecast the value for the next time-period. The forecasted value is simply the average of a fixed number of past values. For time-series data, the choice of the number of fixed values, k, to average is usually a year or a fixed, cyclical period.

Let $x_1, x_2, \ldots, x_n$ denote past time periods with measured/observed values $y_1, y_2, \ldots, y_n$. In this notation, y_n is the current or most recent time period. Using the last k measurements, the predicted value y_{n+1} for the next time period, x_{n+1}, is given by:

$$y_{n+1} = \frac{1}{k} \sum_{j=0}^{k-1} y_{(n-j)} \tag{4}$$

(i.e., a simple average over the last k time periods.)

A more important use of a moving averages technique is to smooth data fluctuations, thereby accentuating a trend. Each data value is replaced by the average of itself and surrounding values. In this way, extreme fluctuations are minimized and an overall trend is displayed.

Computationally, let k be the number of samples on either side of a data value that will be used in averaging. When, for example $k=2$, then 5 values are averaged to give the *averaged* value at y_i, denoted $\bar{y}_i$. In this case,

$$\bar{y}_i = \frac{y_{i-2} + y_{i-1} + y_i + y_{i+1} + y_{i+2}}{5} \tag{5}$$

In general, the new averaged value is computed by:

$$\bar{y}_i = \frac{1}{2k+1} \sum_{j=0}^{2k} y_{[(i-k) + j]} \tag{6}$$

Remark: Each $\bar{y}_i$ in Equations 5 and 6 is a centered-average. Problems will occur at the end points because there will not be enough values to extend out or back k-values for averaging purposes. These points can be dropped if there are a sufficient number of other data points in the sample. Otherwise, a modified average can be computed by repeating the end point values (i.e., extending them) up to k time-periods.

4.3.3 Examples

4.3.3.1 Moving Average: Estimated Reliability of Vehicle Launch. An overall quantitative estimate of the reliability and the trends in reliability of vehicle launch for Atlas/Centaur, Scout, and Delta rockets can be made easily. The estimated reliability, in this context, means the crude ratio of successes to total launches (hence, the probability of success).

Design changes and preflight testing procedures have occurred in the 1960–1988 time period. A simple ratio of total successes to total launches (for each vehicle type) would not reveal the reliability in later years, affected by such changes. Therefore, the data are presented by year. Figures 4.3A through 4.3C give the launch records for: Atlas/Centaur, Delta, and Scout rockets, respectively, in the 1960–1988 time frame.

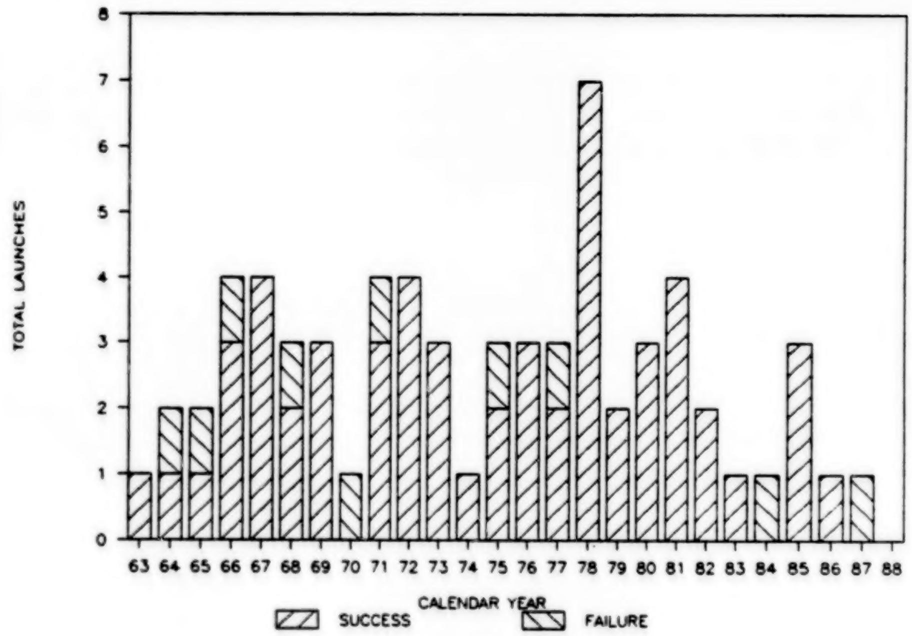


Figure 4.3A. Atlas/Centaur Launch Record

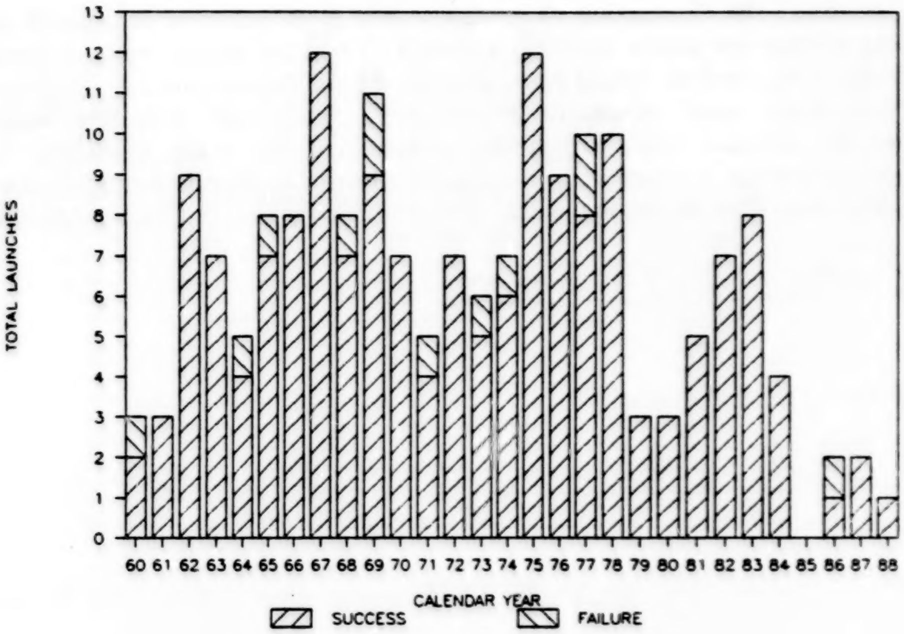


Figure 4.3B. Delta Launch Record

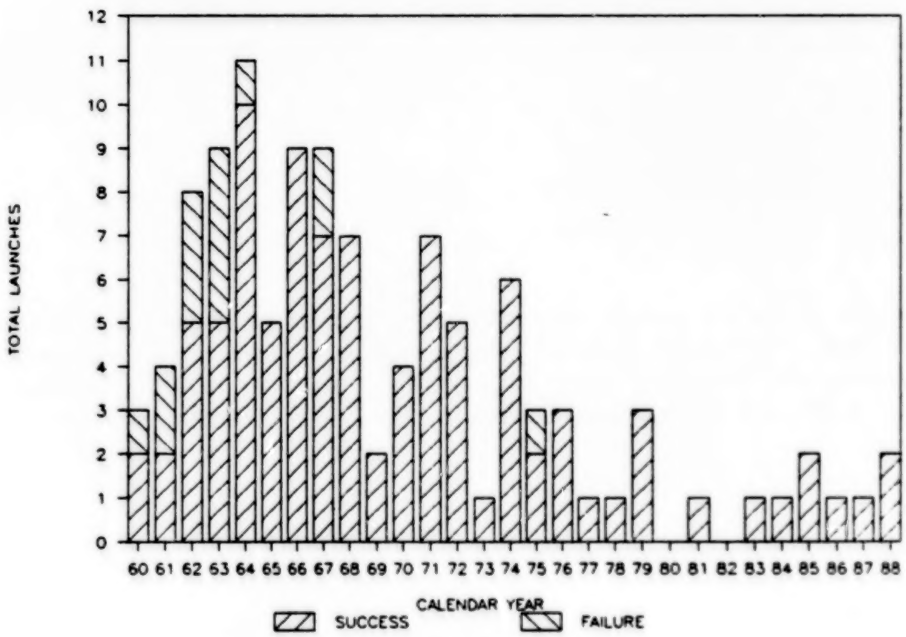


Figure 4.3C. Scout Launch Record

In order to minimize the fluctuation of reliability in a given year due to a small set of launches, a 5-year moving average on yearly reliabilities is used. There are several ways to compute the average yearly reliability. One method would be to average the reliabilities around the given year. However, such a method gives equal weight to the reliability for a year with few launches and to the reliability for an adjacent year, which (by comparison) has many launches. To avoid this, a *weighted* moving average is computed. A weighted moving average is a modification of the moving average method described in section 4.3.2. The average value at y_k, rather than being

$$\bar{y}_k = \frac{y_{k-2} + y_{k-1} + y_k + y_{k+1} + y_{k+2}}{5} \tag{7}$$

is computed as

$$\bar{y}_k = \frac{W_{k-2}Y_{k-2} + W_{k-1}Y_{k-1} + W_kY_k + W_{k+1}Y_{k+1} + W_{k+2}Y_{k+2}}{\Sigma W_i} \tag{8}$$

Here, W_i are the weights, with the standard moving average having all weights equal to 1. Letting

$$\hat{W}_j = \frac{W_j}{\Sigma W_i} \tag{9}$$

note that an equivalent expression for $\bar{y}_k$ is

$$\bar{y}_k = \hat{W}_{k-2}Y_{k-2} + \hat{W}_{k-1}Y_{k-1} + \ldots + \hat{W}_{k+2}Y_{k+2} \tag{10}$$

where

$$\Sigma\hat{W}_j = 1 \tag{11}$$

For the reliability of vehicle launch, the weighting system will be the number of launches attempted in that year. Then, reliabilities determined from a large number of launches will have significant weight and, conversely, reliability values from a small number of launches will have less significant weight. Figure 4.3D gives the weighted moving average launch reliabilities for Atlas/Centaur, Scout, and Delta.

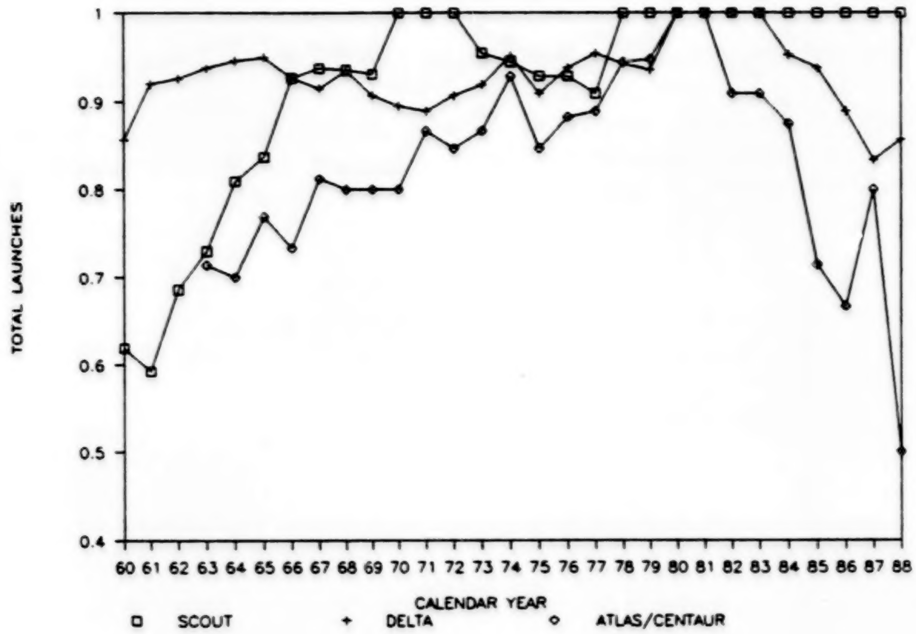


Figure 4.3D. Launch Vehicle Reliability (Five Year Moving Average)

As an example, for a specific computation, the moving average Scout reliability at $t = 1965$ is computed as follows: the reliabilities for 1963 through 1967 are 5/9, 10/11, 5/5, 9/9, and 7/9, respectively. Since there was a total of 43 launches during this period, the weights are 9/43, 11/43, 5/43, 9/43, and 9/43, respectively. Therefore, the computed moving average is given by:

$$\bar{y}_{1965} = 9/43(5/9) + 11/43(10/11) + 5/43(5/5) + 9/43(9/9) + 9/43(7/9)$$
$$= 36/43 = .837 \qquad (12)$$

4.4 FITTING TREND LINES - REGRESSION

4.4.1 Overview and Distinction Between a Trend Line Fit and Regression

4.4.1.1 Overview. Intuitively, suppose at a set of points $x_1, x_2, \ldots, x_n$, we are given a set of measurements/experimental values/observations $y_1, y_2, \ldots, y_n$. One can display these data points (x_i, y_i) graphically as shown in figure 4.4A:

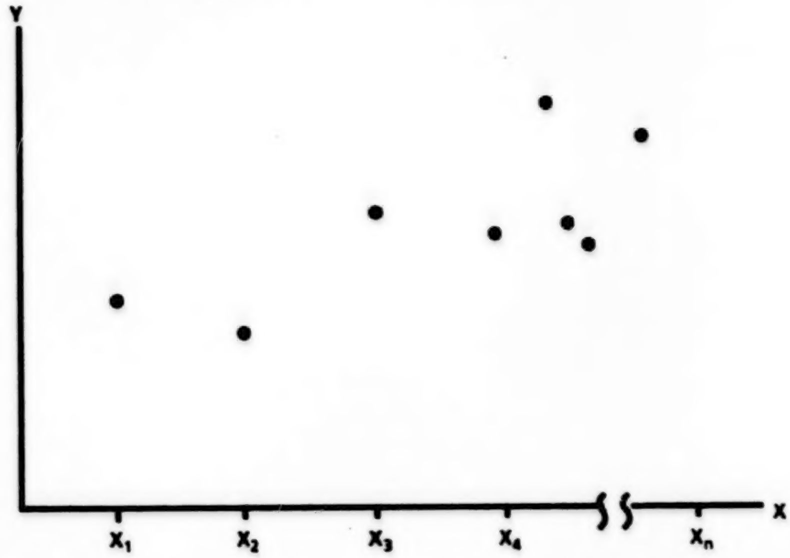


Figure 4.4A. Plot of Rectangular Coordinates $\{(x_i, y_i)\}$

We suspect that there is a relationship between the observation points $\{x_i\}$ and their corresponding observables $\{y_i\}$. The simplest case is when the relationship is linear, so that graphically we have, approximately, a line as shown in figure 4.4B:

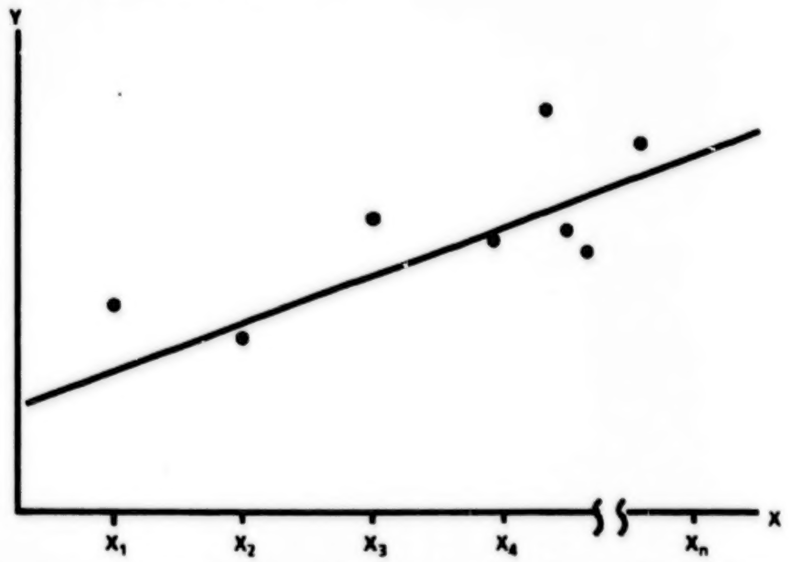


Figure 4.4B. Approximate Linear Relationship Between the Observation Points or Times $\{x_i\}$ and the Observed Values $\{y_i\}$

When the points $\{x_i\}$ are time values, this linear relationship, should it exist, is called a **linear trend**; and the process of finding the *best* line to fit the data is called **fitting or determining a trend line**. Fitting a trend line can be regarded as relatively simple, or somewhat sophisticated, depending upon the point of view. The main subtleties and complexities arise, not from determining the equation of the trend line, but rather in interpreting how accurately this line *fits* the data. In the remainder of this section both the simpler (more practical) viewpoint and a more in-depth presentation are given.

An overview of the process for fitting and evaluating the accuracy of a trend line proceeds as follows:

a. Find the equation of the trend line.

b. Determine whether the trend line, as the best fitting line, adequately models or fits the observed data. If there is a significant lack of fit, then there is no identifiable linear trend.

c. Assuming that the trend line adequately fits the data, test to see if there is a positive or negative (i.e., upward or downward) trend. This requires a statistical test, using what is known as a t-distribution. Briefly, the reason one employs a statistical test is because there is rarely a perfect fit - that is, the $\{y_i\}$ values rarely lie directly on a line. In assuming that a true linear relationship holds between $\{x_i\}$ and $\{y_i\}$, the deviations between each observed y_i and its corresponding point on the trend line are, therefore, considered errors in the measurement of the y_i, or necessary variations in y for a given x_i. These errors, in turn, mean that the actual coefficients of the computed trend line have some error or variation in them. The question then becomes: can one rule out the possibility that the trend line is flat (i.e., no trend), taking into account the errors in estimating the coefficients of the trend line?

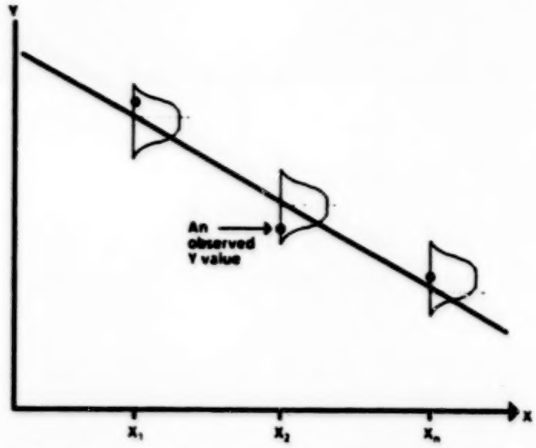
d. Finally, assuming that there is a non-zero trend, use the trend line to predict future values and the *confidence limits* or probable errors surrounding these predicted values.

4.4.1.2 Distinction Between a Trend Line and Regression. Strictly speaking, fitting a trend line is slightly different from the statistical method known as regression analysis. Determining the equation of the best fitting line is the same in both cases; the methods differ in how one interprets how accurately this line *fits* the data, and how one interprets the errors or deviations from the observed values $\{y_i\}$ and the predicted, fitted values $\{\hat{y}_i\}$.

In fitting a trend line, the assumption is usually made that for a given x_i the corresponding y_i is an *exact value*. For example, if one is trending the number of significant problems reported by months prior to launch, the number of significant problems is a constant value for a given month. No error in measurement is postulated. In turn, any deviations between the observed values and the trend line values is attributed solely to lack of fit of a linear model.

The regression analysis viewpoint is more subtle. Note that for each x_i there is a measured or observed value y_i. There could be errors in measuring the value y_i precisely. For example, in a performance trend, if y_i is a continuous variable (e.g., temperature), there will be some variation in measurement. Also, it may be the case that for a given x_i there is not a fixed y_i, but a range of y_i values.

x_i, there is a range of values for y. And that in fitting a line, called the regression line, one is fitting each x_i to the *mean value* of the corresponding observable y. Therefore, at a given x_i there corresponds a distribution of values. The usual assumption is that each measured y_i comes from a normal distribution; and the variance of each distribution, denoted σ^2, is the same (see figure 4.4C).



EACH OBSERVED VALUE IS ASSUMED TO COME FROM A NORMAL DISTRIBUTION CENTERED VERTICALLY AT A GIVEN x_i. THE VARIANCE OF EACH NORMAL DISTRIBUTION IS ASSUMED TO BE THE SAME.

Figure 4.4C. Regression Line

The important point with the regression assumptions is that differences between the observed values $\{y_i\}$ and the corresponding predicted values $\{\hat{y}_i\}$ can be attributed to the necessary randomness (i.e., the distribution) of each y_i. Assuming that a line fits the data, the distributional properties of y at each x_i allow for quantitative statements about the confidence limits, or *bands*, in the estimated slope coefficient, and in values predicted by the fitted line.

4.4.2 Determining a Trend Line by the Least Squares Method

It is evident, from a geometric viewpoint, that there may be many candidates for the best fitting line. For a given line that seems to best approximate the data, a slight perturbation of the line can also be regarded as a good fit. Therefore, a precise notion of the best fitting line is needed.

By definition, the line that best fits the data point $\{x_i, y_i\}$ is that one where the sum of the squares of the deviations between the line at each x_i and y_i is a minimum. Figure 4.4D illustrates this concept.

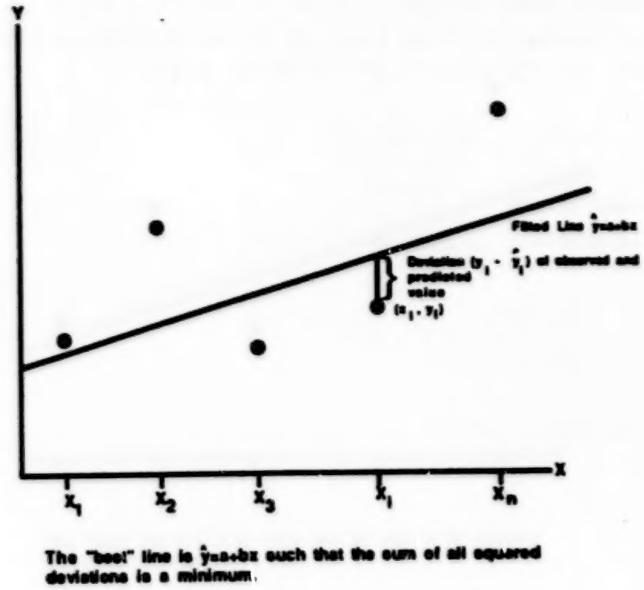


Figure 4.4D. Trend Line (Least Squares Method)

To specify the least squares requirement algebraically, note first that the equation of a line is:

$$y = a + bx \tag{13}$$

where a is the y-intercept and b is the slope.

Thus, the equation of the fitted line will be of form:

$$\hat{y} = a + bx \tag{14}$$

where $\hat{y}$ denotes the *predicted* or *fitted* values.

Requiring that the sum of the square deviations is a minimum, means that

$$\sum_i (y_i - \hat{y}_i)^2 \tag{15}$$

is a minimum, where $\hat{y}_i$ is the value of the fitted line at x_i (i.e., $\hat{y}_i = a + bx_i$).

Since

$$\sum_i (y_i - \hat{y}_i)^2 = \sum_i [y_i - (a + bx_i)]^2 \tag{16}$$

the minimization requirement really means: find the values for a and b to make the sum of squares the smallest. By standard calculus methods (take the partial derivatives with respect to a and b, and set to zero), the solutions for the slope b and intercept a are:

$$b = \frac{n\ \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\ \Sigma x_i^2 - (\Sigma x_i)^2} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}$$

$$a = \frac{\Sigma x_i^2\ \Sigma y_i - \Sigma x_i y_i \Sigma x_i}{n\ \Sigma x_i^2 - (\Sigma x_i)^2} = \bar{y} - b\bar{x} \tag{17}$$

where $\bar{x}$ and $\bar{y}$ denote the means of the $\{x_i\}$ and $\{y_i\}$; i.e.,

$$\bar{x} = \frac{\Sigma x_i}{n} \qquad \bar{y} = \frac{\Sigma y_i}{n} \tag{18}$$

To summarize the above, the **trend line** is that line $\hat{y} = a + bx$, which will minimize the sum of squared deviations between the trend line and the observed values $\{y_i\}$. There is a unique solution for this line, given by:

$$\boxed{\begin{aligned} \hat{y} &= a + bx \qquad \text{(general form)} \\ b &= \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} \\ a &= \bar{y} - b\bar{x} \end{aligned}} \tag{19}$$

Section 4.4.8 gives an example demonstrating the actual calculation of a trend line.

4.4.2.1 Indepth/Theoretical Point-of-View. (*This section is not needed for the basics of obtaining a trend line, and consequently may be bypassed.*) The focus is to present a more rigorous foundation for the least squares method (and hence, trend line fitting). Such ideas give a greater understanding of the least squares method as an **estimation** method. In turn, generalizations beyond a linear model, and questions about the accuracy of fit, are more easily approached and understood.

The least squares principle is an **estimation** method, for which the basic principle is: at the observational points $x_1, x_2, \ldots, x_n$ we have measured values $y_1, y_2, \ldots, y_n$. These measured values may not be the exact true observable values $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$. This could be due to errors in measurements, or the fact that the true, unknown, $\hat{y}_i$ are really **mean** values for a fixed x_i. But we assume that a theoretical model exists, which gives the true $\{\hat{y}_i\}$ given $\{x_i\}$. This model is expressed as a functional dependence:

$$\hat{y} = f(\Theta_1, \Theta_2, \ldots, \Theta_k; x_1, x_2, \ldots, x_n) \tag{20}$$

where the Θ_i are the parameters of the model.

By the least squares method, the best choices (i.e., estimates) for the unknown parameters $\{\Theta_i\}$ are those that minimize the quantity:

$$X^2 = \sum_i [y_i - f(\Theta_1, \Theta_2, \ldots, \Theta_k; x_1, x_2, \ldots, x_n)]^2 \tag{21}$$

In the case of a line (as the model we want to fit),

$$\hat{y} = \Theta_1 + \Theta_2 x \tag{22}$$

Since we want the parameter values $(\Theta_1, \ldots, \Theta_k)$ that minimize X^2, view the Θ_i as variables and minimize by equating all partial derivatives to zero. In the case of fitting a line:

$$\hat{y} = \Theta_1 + \Theta_2 x \tag{23}$$

then

$$X^2 = \sum_{i=1}^{n} (y_i - \hat{y})^2 = \sum_{i=1}^{n} (y_i - \Theta_1 - \Theta_2 x_i)^2 \tag{24}$$

Then

$$\begin{aligned} \frac{\partial X^2}{\partial \Theta_1} &= \sum_{i=1}^{n} (-2)(y_i - \Theta_1 - \Theta_2 x_i) = 0 \\ \frac{\partial X^2}{\partial \Theta_2} &= \sum_{i=1}^{n} -2x_i (y_i - \Theta_1 - \Theta_2 x_i) = 0 \end{aligned} \tag{25}$$

Solving for the parameters Θ_1, and Θ_2 gives:

$$\begin{aligned} \hat{\Theta}_1 &= \frac{\Sigma x_i^2 \, \Sigma y_i - \Sigma x_i y_i \, \Sigma x_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \\ \hat{\Theta}_2 &= \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \end{aligned} \tag{26}$$

Strictly speaking, note that $\hat{\Theta}_1$, and $\hat{\Theta}_2$ are functions solely of the observations (x_i, y_i). Thus, $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are **estimators** of the true, but unknown, parameters Θ_1 and Θ_2. This means that, for each set of $\{(x_i, y_i)\}$ observations, from the estimators $\hat{\Theta}_1$, $\hat{\Theta}_2$ one can compute an estimated value of each true parameter Θ_1 and Θ_2. In general, the distribution of these estimates about the true parameter values have certain optimal properties called unbiasedness and minimum variance.

The final point is that, with $\hat{y} = \Theta_1 + \Theta_2 x$, the equation is **linear** *in the parameters.* In general, when trying to estimate the coefficients (i.e., parameters) of a higher order polynomial, $\hat{y} = \Theta_1 + \Theta_2 x + \ldots + \Theta_{k+1} x^k$, the parameter dependence is also linear, even though the equation is not linear in x. It will follow that the same methods used in estimating the parameters for a line

apply, and exact solutions exist. Thus, fitting a parabola or higher order polynomial can be viewed as a *linear* regression or estimation, as opposed to a non-linear problem.

In summary, the least squares method is a technique for estimating parameter values of a model. One assumes that there is a functional model of observational points $\{x_i\}$ and unknown parameters that predicts the true values associated with each x_i. By minimizing the sums of squared deviations (least squares principle), estimators $(\hat{\Theta}_1,\ldots,\hat{\Theta}_k)$ of the true parameters $(\Theta_1,\ldots,\Theta_k)$ are obtained. For a given set of measurements $\{(x_i, y_i)\}$, using the estimators will give an *estimated value of the parameter*. These estimated values are then used in the function $f(\Theta_1, \Theta_2,\ldots,\Theta_k; x_1,x_2,\ldots,x_n)$ to obtain a predicted value $\hat{y}$ for a given set of observational points $x_1,x_2,\ldots,x_n$.

4.4.3 Determining Whether the Estimated Trend Line Fits the Data - Use of R^2 and χ^2 Goodness-of-Fit Test

4.4.3.1 Intuition. Having determined the equation of the trend line, the next issue involves the degree to which a line adequately describes the data. Put another way, how good a fit is the line? To clarify the question of fit, figure 4.4E gives two different sets of data; both sets, however, have the *same* least-squares-derived trend line.

(a) (b)

Figure 4.4E. Trend Line — Fit of Sample Data

It is intuitively clear that the line in figure 4.4E(a) describes the data almost exactly. In turn, qualitatively, there is a high degree of fit. With reasonable confidence, the estimated trend line can be used to predict future values based on future time (i.e., larger x) values. In contrast, the data points in figure 4.4E(b) move up and down around the trend line, with large deviations. Intuitively, the trend line does not adequately model the data, and there is little confidence in prediction of future values.

4.4.3.2 R^2 as a Measure of Fit. The preceding section and figure show that a *quantitative measure* is needed to measure the precision, or accuracy, of a fitted trend line. One such measure is

called the **R-square value**, denoted R^2. This measure is also known as the **coefficient of determination.**

To motivate the derivation of R^2, let $\hat{y} = a + bx$ denote the fitted line. An obvious measure of how well $\hat{y}$ matches the observed y_i values is to look at the *difference* between the observed and predicted values: $\epsilon_i = y_i - \hat{y}_i$. If there is a close fit, these differences, called **residuals**, are small. Since only the magnitude, not the sign of each ϵ_i is important, we consider $\epsilon_i^2 = (y_i - \hat{y}_i)^2$; and as a measure of total fit, sum up the squared residuals, $\Sigma\epsilon_i^2$.

Now, a small (large) $\Sigma\epsilon_i^2$ would indicate a good (poor) fit. The problem with this measure is that there is no relative scale. That is, if the y_i are large values, and $\hat{y}$ fits well, then ϵ_i could still be *large*, in contrast to the case when the y_i are small, and ϵ_i is still comparatively large relative to y_i.

The way to normalize and evaluate the magnitudes of the ϵ_i relative to their corresponding y_i, is to divide $\Sigma\epsilon_i^2$ by the quantity $\Sigma(y_i-\bar{y})^2$. This quantity is called the **variation about the mean**, and is the sum of squared deviations about the constant line $y = \bar{y}$, where $\bar{y}$ denotes the mean of $\{y_i\}$.

Consider the identity:

$$(y_i-\bar{y}) = (y_i-\hat{y}_i) + (\hat{y}_i-\bar{y}) \tag{27}$$

Squaring both sides, and summing over the n observations $i = 1, 2, \ldots, n$, we get:

$$\Sigma\,(y_i-\bar{y})^2 = \Sigma\,(y_i-\hat{y}_i)^2 + \Sigma\,(\hat{y}_i-\bar{y})^2 \tag{28}$$

The cross-product term $2\Sigma(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)$ can be shown to vanish.

The measure, R^2, follows immediately from the above identity. Note that the variation about the mean, $\Sigma(y_i-\bar{y})^2$, has been partitioned into two sums of squares — the sum of residuals squared $\Sigma(y_i-\hat{y}_i)^2$ and the variation of the regression line about the mean, $\Sigma(\hat{y}_i-\bar{y})^2$.

By definition,

$$R^2 = \frac{\Sigma\,(\hat{y}_i-\bar{y})^2}{\Sigma\,(y_i-\bar{y})^2} \tag{29}$$

It follows from equation (28) that when R^2 is close to 1, then $\Sigma\epsilon_i^2 = \Sigma(y_i - \hat{y}_i)^2$ is close to zero. This is the desired criterion, since then the deviations between the observed y_i and the fitted values, $\hat{y}_i$, are small; i.e., there is a good fit. In contrast, when R^2 is close to 0, most of the variation about the mean (a fixed quantity) is not explained by the fitted line, but to the fact that most y_i do not lie close to the fitted trend line.

To summarize:

> R^2 (the coefficient of determination) measures how accurately the trend line fits the data.
>
> $$R^2 = \frac{\Sigma\ (\hat{y}_i-\bar{y})^2}{\Sigma\ (y_i-\bar{y})^2},\quad 0 \le R^2 \le 1 \qquad (30)$$
>
> R^2 close to 1 means there is a better fit

Remark: The quantity R^2 derives from the identity $\Sigma\ (y_i-\bar{y})^2 = \Sigma\ (y_i-\hat{y}_i)^2 + \Sigma\ (\hat{y}_i-\bar{y})^2$ by dividing both sides by $\Sigma\ (y_i-\bar{y})^2$. Thus,

$$1 = \frac{\Sigma\ (y_i-\hat{y}_i)^2}{\Sigma\ (y_i-\bar{y})^2} + R^2 \qquad (31)$$

An equivalent expression for R^2 is:

$$R^2 = 1 - \frac{\Sigma\ \epsilon_i^2}{\Sigma\ (y_i-\bar{y})^2} \qquad (32)$$

4.4.3.3 Minimum Values of R^2 for Concluding that the Trend Fit is Significant - The Distribution of R^2. The R^2 statistic gives a measure of how well the trend model fits the data. With regard to quantification of fit, we have previously noted that R^2 ranges between 0 and 1, with 1 indicating a perfect fit.

A more precise quantititative criterion for fit using R^2 can be established. Such a criterion uses distributional assumptions for the error terms, $\epsilon_i = y_i - \hat{y}_i$ and for the form of the model. The following table gives the minimum values for R^2 to conclude that the fitted model is an adequate description of the data. The table is based on the number of sample points and confidence level (i.e., 100 $(1-\alpha)$% where α is the significance level). It is important to note that the table is valid only for two-parameter models; i.e., models with 2 unknown variables a and b such as in $y = a + bx$ or $y = ae^{bx}$.

MINIMUM R-SQUARE VALUES FOR SIGNIFICANT TRENDING FIT

NUMBER OF DATA POINTS (N)	$\alpha = .01$	$\alpha = .025$	$\alpha = .05$
4	.98	.95	.90
5	.92	.85	.77
6	.84	.75	.66
7	.76	.67	.57
8	.70	.59	.50
9	.64	.54	.44
10	.59	.49	.40
11	.54	.44	.36
12	.50	.41	.33
13	.47	.38	.31
14	.44	.35	.28
15	.41	.33	.26
20	.31	.25	.20
25	.26	.20	.16
30	.21	.17	.13

To use the above table, locate the number of observational values (N) and the desired level of significance (almost always use $\alpha = .05$). The corresponding table entry gives the *minimum* R^2 value to conclude that the fit is significant. For example, where $N = 8$, then R^2 should be at least .5 (here $\alpha = .05$) to conclude that the trend model is an adequate description of the data.

4.4.4 Testing for a Non-Zero Trend — t-Test on the Slope Coefficient

Assume that a trend line has been fit to the data. Therefore, one has the best linear approximation or fit:

$$\hat{y} = a + bx \tag{33}$$

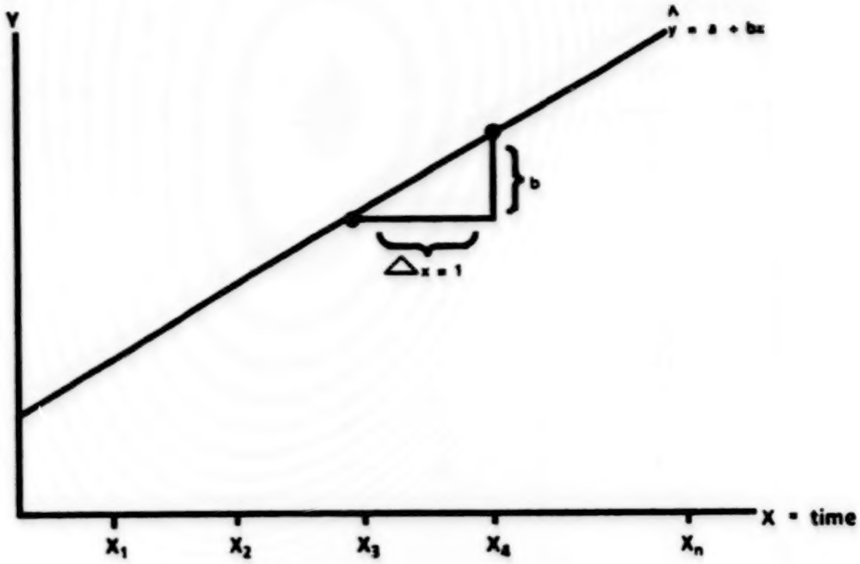
where a and b are constants that have been determined by the least squares method.

It is important to note that we assume that there is a reasonable degree, or accuracy, of fit. This determination could follow from inspection of the data, the R^2 value, or other methods noted in section 4.4.3. If a linear model (i.e., a line) suffers from lack of fit, there is no point in proceeding further to determine quantitatively whether there is a linear trend.

With the above assumptions, the next question is: does the fitted trend line imply a non-zero trend; i.e., is there a statistically significant upward or downward movement exhibited over time? This question has both a simple and a more complex answer.

In the less rigorous, simpler approach, note that the constant b gives the slope of the trend line. Assuming the model is correct (i.e., an accurate fit), then when $b = 0$, there is no trend; and for

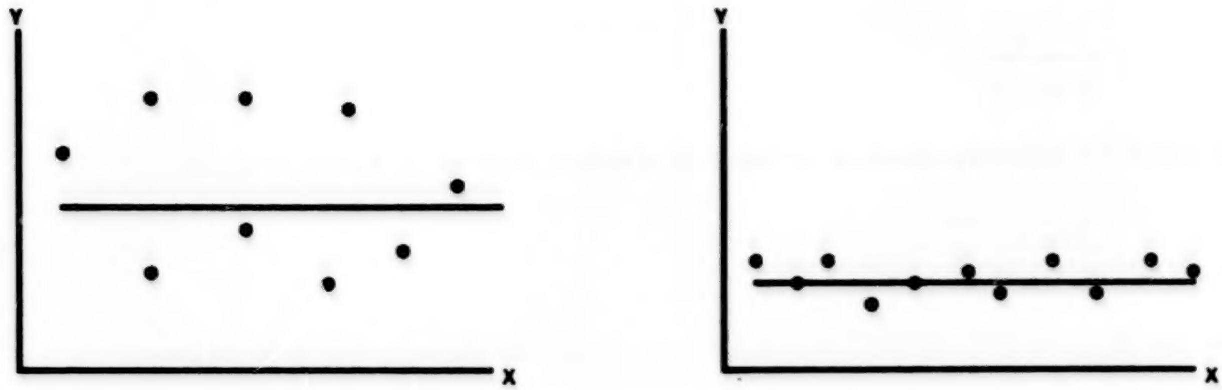
b>0 (b<0) there is an increasing (decreasing) trend over time. For each unit of time, the amount of increase (decrease) in the variable y is b. Graphically, this is depicted in figure 4.4F.



FOR EACH UNIT OF TIME x, THERE IS A CONSTANT INCREASE OF b UNITS IN y.

Figure 4.4F. A Non-Zero Increasing Trend

Prior to a more indepth approach to testing for trend, an important distinction or clarification should be made. Concerning the meaning of the statement: the data does not indicate a (linear) trend; or there is *no trend*. This statement has two distinct meanings. Figure 4.4G graphically illustrates the distinction. For both cases, no trend exists, but the reasons are entirely different. In figure 4.4G(a), no trend is exhibited because there is no adequate fit of a trend line to the data. In 4.4G(b), there is also *no trend* or zero-trend, but here a linear model does adequately fit the data; however, the line is flat (i.e., has 0 slope).



(a) NO TREND DUE TO LACK-OF-FIT

(b) MODEL FITS; BUT NO TREND BECAUSE OF 0-SLOPE TREND LINE

Figure 4.4G. Two Examples of Zero-Trend Lines

4.4.4.1 More Rigorous Basis in Testing for Non-Zero Trend. In determining the trend line coefficients a and b, no assumptions were made concerning the distributional properties of the observations $\{y_i\}$. To determine how reliable these estimated coefficiencies are, assumptions are made about the probability distributions of each y_i, or equivalently about each error term $\epsilon_i = y_i - \hat{y}_i$. The aim is to determine certain *bands* or *confidence limits* around the estimated trend line slope b - to judge whether one can conclude that there is a statistically significant non-zero trend. Put another way, is b, in all probability, different from zero?

More rigorously, one needs to determine the standard deviation of the slope b, and confidence intervals. The standard assumption is that the observables, y_i, are each normally distributed about this mean $\hat{y}_i$ and that the y_i are independent. These conditions are equivalent to requiring that the error terms (also called residuals) are normally distributed with mean 0. Thus, in the linear model:

$$y_i = a + bx_i + \epsilon_i \qquad \text{for } i = 1, 2, \ldots, n \tag{34}$$

These assumptions are shown geometrically in figure 4.4C.

Based on the distribution assumptions for the measured values y_i (or equivalently the ϵ_i), it can be shown that 90% confidence intervals for b are given by:

$$b \pm \frac{t_{(n-2,\ .95)} \cdot s}{[\Sigma\ (x_i - \bar{x})^2]^{1/2}} \tag{35}$$

where s is the estimated standard deviation of the residuals. The value $t_{(n-2,\ .95)}$ is the 95% point of a t-distribution with (n-2) degrees of freedom.

The **estimated standard deviation** of ϵ_i denoted s, is given by:

$$s = \left[\frac{\Sigma\ (y_i-\hat{y}_i)^2}{n-2} \right]^{1/2} \qquad (36)$$

and the quantity

$$\frac{s}{[\Sigma\ (x_i-\bar{x})^2]^{1/2}} \qquad (37)$$

is called the **estimated standard deviation** or **standard error** of b. Letting γ denote

$$\frac{t_{(n-2,\ .95)} \cdot s}{[\Sigma\ (x_i-\bar{x})^2]^{1/2}} \qquad (38)$$

we can say with 90% confidence that the true trend line slope, β, lies in the interval

$$[b - \gamma,\ b + \gamma] \qquad (39)$$

Note that if the above interval does not include 0, then there is a statistically significant trend (at the 90% confidence level). Hence, there is a *non-zero trend*. When the interval lies completely to the right of 0, there is a 90% confidence that the true slope is positive. Therefore, there is an increasing trend. For an interval to the left of 0, there is a 90% confidence that the slope is negative, and therefore a decreasing trend. The key point here is that having the slope estimate b, we can construct an interval $[b - \gamma, b + \gamma]$, for which we have a 90% confidence that the true (but unknown) slope lies in this interval. Based on the above discussion, we cannot attach any confidence level to what the actual value of the true slope is. The following test gives a method for testing whether the true slope is equal to zero (hence no trend exists).

An equivalent way to test whether b is significantly different from zero (and hence, a non-zero trend) is as follows:

the variable $t = \dfrac{b[\Sigma(x_i-\bar{x})^2]^{1/2}}{s}$ has a t-distribution.

When there are n data points, there are (n-2) *degrees of freedom*. This number, n-2, characterizes the distribution completely.

First compute t for the data sample in question. Next locate $|t|$ (or the *nearest value* $\leq |t|$) in the statistical table of t-values for n-2 degrees of freedom. (See Table 4.4-1 at the end of this section.) The column that $|t|$ appears in (labeled *F* in the table) gives the percentage (divided by 100) of the distribution that would have this t-value or less - while having b equal to 0. This number is labeled $1-\alpha$. Therefore, α gives the probability to get the computed t-value or greater *and still have no trend*. (Technically since we are considering $|t|$, the actual probability is $\alpha/2$.) Thus, when α is small (say .025), there is little chance of getting the computed t-value without a positive or negative trend. For $\alpha \leq .025$, we conclude that the true value of the slope is, in fact, different from zero; hence there is a trend. Note that, in general, the higher the t-value, the greater the likelihood that the trend slope is different from zero.

To further clarify the above procedure, suppose that there are 10 data points and the computed value of t is 2.15. Then, there are 8 degrees of freedom (i.e., n-2, when n = 10). This computed t-value is between the table t-values of 1.860 and 2.306. The percentage point of the distribution for a t-value of 1.86 is .95. Therefore, there is only a 5% chance of getting a t-value of 1.86 or greater (and hence of 2.15) while having no trend. Since we are considering both t and -t, more technically, the actual probability is .05/2 = .025 or 2.5%.

The above test is used to test the hypothesis that the true trend slope is zero. When the |t| is large (i.e., exceeds the 95th percent point), we can conclude that the true slope, b, is not zero. However, when the t value does not exceed the 95% point, we cannot rule out the possibility that the true value is zero. (This does not mean, however, that we accept the hypothesis that the value is zero—only that we cannot rule it out).

This test can be generalized easily to provide a method for testing the hypothesis that the slope is equal to any number b_0. Such a test would be necessary to ascertain whether the trend slope could be a certain, specific value. Recall that we have an estimated slope b (derived from the least squares method). Analogously to the variable

$$t = \frac{b[\Sigma\ (x_i-\bar{x})^2]^{1/2}}{s} \tag{40}$$

having a t-distribution, we have that the variable

$$t = \frac{(b-b_0)[\Sigma\ (x_i-\bar{x})^2]^{1/2}}{s} \tag{41}$$

also has a t-distribution. Using the same procedure as given above, one computes t and locates |t| in the table. For the corresponding percentage point, denoted $1-\alpha$, we have that α is the probability of obtaining this |t| value (or higher) and having the true slope be b_0. When α is small (less than .05), we conclude that there is less than a 5% chance that the true slope could be b_0.

For practical purposes, whenever the probability of no-trend is 10% or less with the computed t-value, we will assume that there is a trend (i.e., $b \mp 0$). Consequently, to test for a non-zero trend the following procedure can be used:

(1) Compute t-value.
(2) When there are n data points, there are n-2 degrees of freedom.
(3) Read down the .95 column of the t-distribution table to the row corresponding to n-2 degrees of freedom.
(4) If the computed t-value is **greater than** or **equal to** the located table value, there is a significant non-zero trend.

4.4.5 Prediction of Values Using the Trend Line

The fitted trend line is given by:

$$\hat{y} = a + bx \qquad (4\ )$$

However, both estimated coefficients, a and b, are subject to error. This error, in turn, will influence the predicted value variable, $\hat{y}$. If we ignore these errors, then the prediction of $\hat{y}$ for a given value of x is straightforward. By simple substitution, when x equals x_0, then $\hat{y}_0 = a + bx_0$. Usually x_0 represents a future time for prediction purposes.

Simple substitution in the trend line equation gives a predicted value. To ascribe confidence limits to this predicted value, the standard deviation of $\hat{y}_0$ at a given x_0 is needed. It can be shown that the estimated standard deviate of $\hat{y}_0$ is:

$$\text{estimated standard deviation } (\hat{y}_0) = s\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\Sigma\,(x_i-\bar{x})^2}\right]^{1/2} \qquad (43)$$

where,

$$s^2 = \frac{\Sigma\,(y_i-\hat{y}_i)^2}{n-2} \qquad (44)$$

The 90% confidence limits for a predicted value $\hat{y}_0$ are:

$$\hat{y}_0 \pm t_{(n-2,\ .95)} \cdot s\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}\right]^{1/2} \qquad (45)$$

Here, $t_{(n-2,\ .95)}$ is the t-value corresponding to the 95th percentage point with n-2 degrees of freedom.

Table 4.4-1

PERCENTAGE POINTS OF THE STUDENT'S t-DISTRIBUTION

The table gives values of t_α for different degrees of freedom υ such as to produce specified values $F(t_\alpha;\upsilon)$, where

$$F(t_\alpha;\upsilon) = \int_{-\infty}^{t_\alpha} \frac{\Gamma[\frac{1}{2}(\upsilon+1)]}{\sqrt{\pi\upsilon}\;\Gamma(\frac{1}{2}\upsilon)} \frac{1}{\left[1+\frac{t^2}{\upsilon}\right]^{\frac{1}{2}(\upsilon+1)}}\, dt = 1-\alpha$$

$F(-t_\alpha;\upsilon) = 1 - F(t_\alpha;\upsilon)$, $\upsilon = \infty$ corresponds to the standard normal distribution.

υ \ F	.60	.70	.80	.90	.95	.975	.990	.995	.999	.9995
1	.325	.727	1.376	3.078	6.314	12.706	31.821	63.657	318.31	663.62
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	.257	.532	.859	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	.256	.532	.858	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	.256	.532	.858	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	.256	.531	.857	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	.256	.531	.856	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	.256	.531	.856	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	.256	.531	.855	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	.256	.530	.855	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	.256	.530	.854	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	.256	.530	.854	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	.255	.529	.851	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	.255	.528	.849	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	.254	.527	.848	1.296	1.671	2.000	2.390	2.660	3.232	3.460
70	.254	.527	.847	1.294	1.667	1.994	2.381	2.648	3.211	3.435
80	.254	.527	.846	1.292	1.664	1.990	2.374	2.639	3.195	3.416
90	.254	.526	.846	1.291	1.662	1.987	2.369	2.632	3.183	3.402
100	.254	.526	.845	1.290	1.660	1.984	2.364	2.626	3.174	3.391
110	.254	.526	.845	1.289	1.659	1.982	2.361	2.621	3.166	3.381
120	.254	.526	.845	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	.253	.524	.842	1.282	1.645	1.960	2.326	2.576	3.090	3.291

The *90% confidence limit* means that there is a .9 probability that the true value of y at x_0 will lie in the interval

$$[\hat{y}_0-\gamma,\ \hat{y}_0+\gamma] \tag{46}$$

where,

$$\gamma = t_{(n-2,\ .95)} \cdot s \cdot \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\Sigma(x_i-\bar{x})^2} \right]^{1/2} \tag{47}$$

To summarize:

> The trend line equation is
>
> $\hat{y} = a + bx,$
>
> where $\hat{y}$ is the predicted value.
>
> The predicted value $\hat{y}_0$ at a point x_0 is found by simple substitution.
>
> Confidence limits around the predicted value can be computed - see text and figure 4.4H. These limits or bands represent the most reasonable range for the true value of y at x_0.

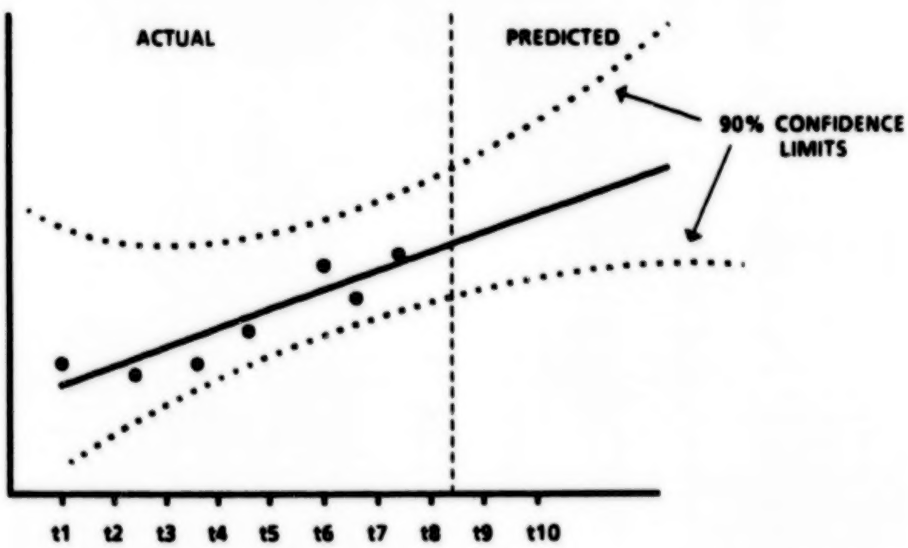


Figure 4.4H. Fitted Trend Line and Extrapolation of Trend Line to Future Time Values (with 90% Confidence Limits on Predicted Values)

4.4.6 Extension of Linear (Line) Fit to Polynomial and Exponential Cases

The linear (line) model represents the simplest trend. However, data may be described more aptly by a quadratic (polynomial of degree 2) or exponential model. These models are given by:

$$\begin{aligned} y &= a_0 + a_1 x + a_2 x^2 \qquad &\text{(quadratic)} \\ y &= ae^{bx} \qquad &\text{(exponential)} \end{aligned} \tag{48}$$

Graphically, these models produce the curves shown in figure 4.41.

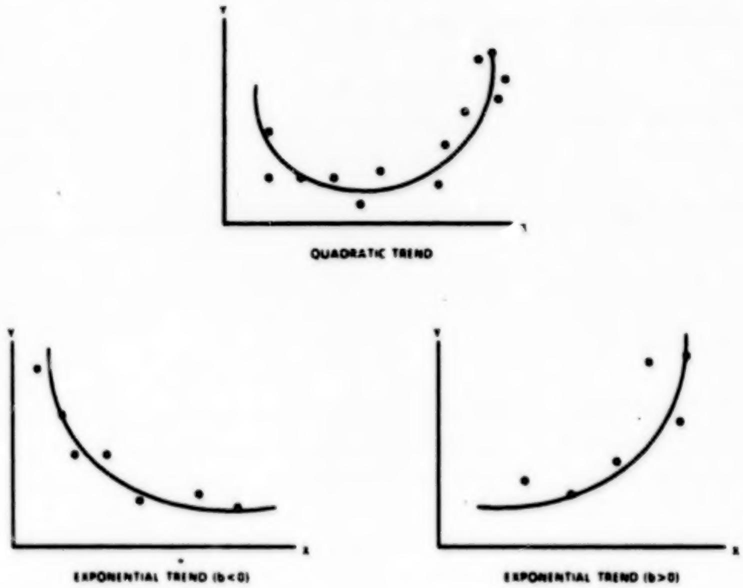


Figure 4.41. Quadratic and Exponential Models

With regard to trending, a **linear trend** means that the variable of interest y increases (decreases) by a *constant* amount for each unit of time. Hence, the rate of change of y is constant:

$$\left[\frac{dy}{dt} = k \right] \tag{49}$$

In a **quadratic trend**, y increases (decreases) by an amount that is proportional to the time. Thus, the rate of increase (decrease) is *not constant*; but, the rate of change of the rate itself is *constant*:

$$\left[\frac{dy}{dt} = kt; \ \frac{d}{dt}\left(\frac{dy}{dt}\right) = k \right] \tag{50}$$

With an **exponential trend** the amount of increase (decrease) in y per unit time is proportional to the magnitude of y:

$$\left[\frac{dy}{dt} = ky \right] \tag{51}$$

Therefore, the *percentage of change per unit time is constant*:

$$\left[\frac{dy/dt}{y} = k \right] \tag{52}$$

Overall, then, the progression of trend models is:

linear (increase/decrease is constant)	→ quadratic (increase/decrease is linear)	→ exponential (% increase/decrease is constant)

4.4.6.1 Polynomials of Order ≥ 2. A polynomial of order n has an equation of form:

$$f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \tag{53}$$

where the a_i are constants.

Seldom, in a trending examination, is it useful to consider any polynomial of degree greater than 2. Polynomials of higher degree ($n \geq 3$) usually have local maxima and minima; and, hence, are not very useful for discerning global upward or downward movements. Additionally, use of higher degree polynomials can be misleading, because given $(n+1)$ data points, a polynomial of degree n can be found that exactly fits the data (e.g., if there are 4 y-values corresponding to 4 distinct times x_1 through x_4, then a cubic polynomial will exactly fit the data).

For the above reasons, polynomial trending fitting should usually be limited to second-degree (quadratic) polynomials. However, the methods described below for quadratics are equally applicable for polynomials of greater degree.

A quadratic polynomial has an equation of form:

$$f(x) = a_0 + a_1x + a_2x^2 \tag{54}$$

The least squares method can be used to determine the best estimates for coefficients a_0, a_1, and a_2. Strictly speaking the polnomial is *linear in the parameters* to be estimated (the a_i); therefore, the estimation process is a linear regression (not a non-linear regression problem).

Notationally, the fitted quadratic is written:

$$\hat{y} = a_0 + a_1x + a_2x^2 \tag{55}$$

where $\hat{y}$ is the predicted value for a given x, and the a_i are the estimated coefficients.

The least squares estimation method yields 3 equations in the 3 unknown parameters a_0, a_1, a_2. These equations are solved simultaneously for a_0, a_1, a_2. The equations are:

$$\boxed{\begin{aligned} \Sigma y_i &= a_0 n + a_1 \Sigma x_i + a_2 \Sigma x_i^2 \\ \Sigma y_i x_i &= a_0 \Sigma x_i + a_1 \Sigma x_i^2 + a_2 \Sigma x_i^3 \\ \Sigma y_i x_i^2 &= a_0 \Sigma x_i^2 + a_1 \Sigma x_i^3 + a_2 \Sigma x_i^4 \end{aligned}} \tag{56}$$

4.4.6.2 Exponential Trend Model. An exponential model is of form:

$$f(x) = ae^{bx} \tag{57}$$

where a and b are constants.

In a trending context, when $b>0$, the model will approximate a trend where the response variable y *rapidly increases* in relation to increasing values of x; (when $b<0$, the response variable y *rapidly decreases* in relation to increasing values of x).

The above equation is non-linear in the parameter, b; however, the equation can be transformed into an equation that is linear in b and in ln a; thus, the parameter b (and a) can be estimated using the least squares method. Writing the predicted or *fitted* value as $\hat{y}$ and taking the natural log yields:

$$\begin{aligned} \hat{y} &= ae^{bx} \quad \Rightarrow \\ \ln \hat{y} &= \ln a + bx \end{aligned} \tag{58}$$

Regard ln a as a new variable, $\tilde{a}$ to give:

$$\ln \hat{y} = \tilde{a} + bx \tag{59}$$

Since the observed value $\{y_i\}$ are to be fitted by $\{\hat{y}_i\}$, it follows that $\ln y_i$ will be modeled by $\ln \hat{y}_i = \tilde{a} + bx_i$. The least squares method yields two equations in the two unknowns $\tilde{a}$ and b:

$$\begin{aligned} \Sigma \ln y_i &= \tilde{a} n + b \Sigma x_i \\ \Sigma x_i \ln y_i &= \tilde{a} \Sigma x_i + b \Sigma x_i^2 \end{aligned} \tag{60}$$

Solving these equations simultaneously gives exact solutions for $\tilde{a}$ and b. The estimate for a is found from $\tilde{a}$ by transforming back:

$$\begin{aligned} \tilde{a} &= \ln a \quad \Rightarrow \\ e^{\tilde{a}} &= a \end{aligned} \tag{61}$$

Having the least squares estimates, a and b, the best fitting exponential trend equation, $\hat{y} = ae^{bx}$, is then determined by simple substitution.

4.4.6.3 Basic Trend Models. Based on past experience in trending Shuttle hardware failure modes and hardware performance, the most useful trend models are:

Linear:	$f(x) = a + bx$
Exponential:	$f(x) = ae^{bx}$ or $\ln(f(x)) = \ln(a) + bx$
Logarithmic:	$f(x) = a + b\ln(x)$
Power:	$f(x) = ax^b$ or $\ln(f(x)) = \ln(a) + b\ln(x)$
Quadratic:	$f(x) = (a + bx)^2$

4.4.7 Applicability

The fitting of lines, polynomials of order ≥ 2, and exponentials are basic techniques in trend analysis. The methods described in the preceding sections apply to all types of trending: programmatic, performance, problem, and supportability. In particular, the methods discussed should be used for performance trending, since the variables measured are continuous and usually have an uncertainty in measurement.

To determine which trending model, if any, is most applicable, the suggested first step is to plot the data. Inspection will often suggest the most reasonable analytic model. A quantifiable method to determine whether a linear, polynomial of degree 2, or exponential model is more appropriate is as follows:

The more closely a model fits the observed data, the smaller will be the squared deviations between the predicted and observed values. Therefore, the *smaller* the quantity

$$\frac{\Sigma\,(y_i-\hat{y}_i)^2}{n - L} \tag{62}$$

called the **estimated variance about regression**, the closer the fit.

The constant L equals the number of unknown parameters in the model. The model with the smallest estimated variance about regression should be used.

4.4.8 Examples

4.4.8.1 Programmatic Trends on Orbiter/Orbiter Tile Wad Closures. Figures 4.4J and 4.4K display the number of Work Authorization Documents (WADs) at KSC in the weeks prior to shuttle launch. The tile WADS and the remaining orbiter directives are followed separately. There should be a declining number (hence decreasing trend) in open WADs as the time-to-launch approaches. A display of the magnitude of open WADs and the rate of decrease provide decision-makers with

information on workloads, scheduling, projected overtime, manpower resource allocations, and general suitability of launch date.

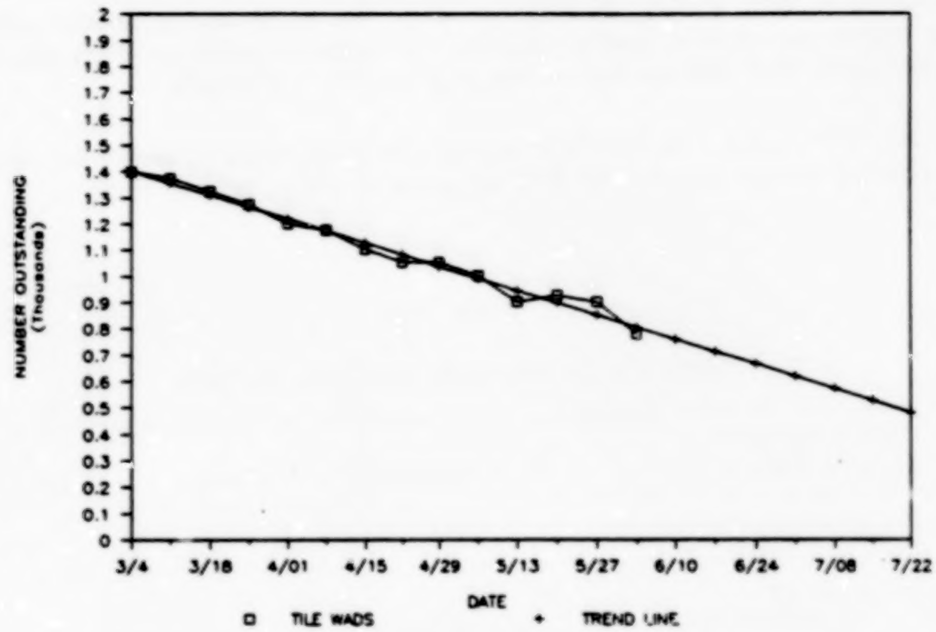


Figure 4.4J. Trend of Tile WAD Closure (as of 6/03)

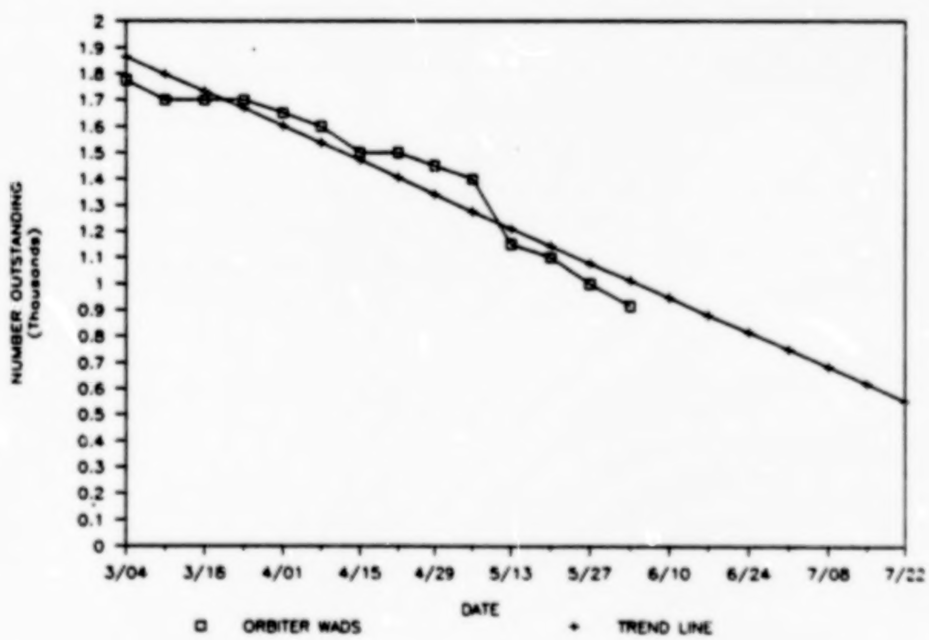


Figure 4.4K. Trend of Orbiter WAD Closure (as of 6/03)

4.4.8.2 Multiplexer/Demultiplexer (MDM): Problem Trends in Design, Manufacturing, and Electrical/Electronic/Electromechanical (EEE) Failures. Multiplexer/demultiplexers in the Space Shuttle Orbiter act as data acquisition, distribution, and signal conditioning units. The units accept digital information from their controlling units (IOP, PCM, or GSE) and convert or reformat the digitized information into analog, discrete, or series digital signals for shuttle subsystems. Conversely, the MDMs can receive analog, discrete, or digital information from shuttle subsystems, digitize these signals, and transfer this information to their controllers.

There have been 776 failures in the 1976 through 1986 timeframe. Table 4.4-2 summarizes these failures by manufacturing, design, and EEE in terms of where failures occurred.

Table 4.4-2

MDM FAILURES FOR 1976-1986 PERIOD

Where Occurred (No. of Failures)	Manufacturing	Design	EEE Parts
Acceptance (442)	53	60	329
Certification (12)	0	6	6
Field (318)	8	69	241
Flight (4)	0	2	2
TOTALS (776)	61	137	578

The following graphs (figures 4.4L through 4.4O) depict the frequencies and trends in manufacturing, design, and EEE failures during the 1976 to 1986 timeframe. (The rate of MDM manufacturer acceptance and field tests are essentially constant; hence, no normalization is necessary.)

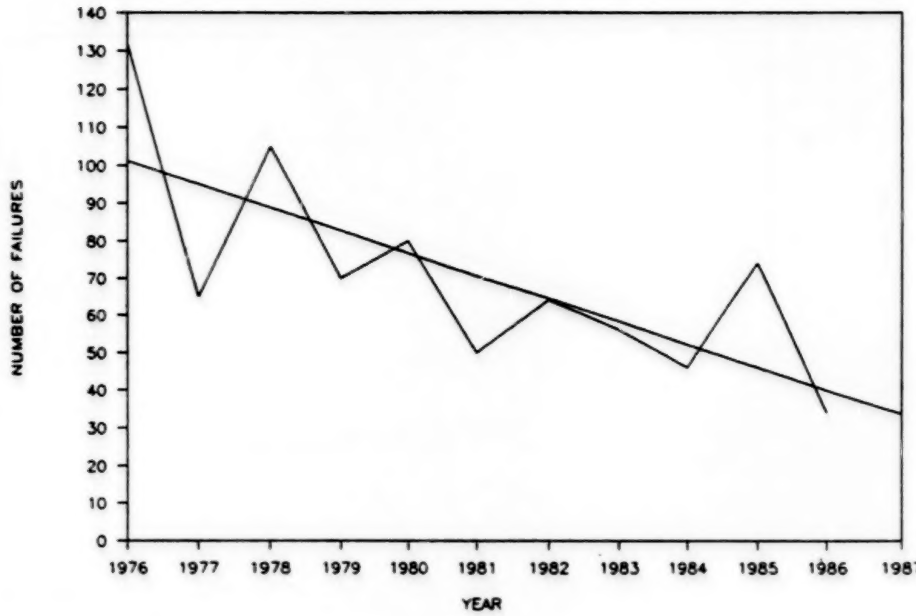


Figure 4.4L. Total MDM Failures

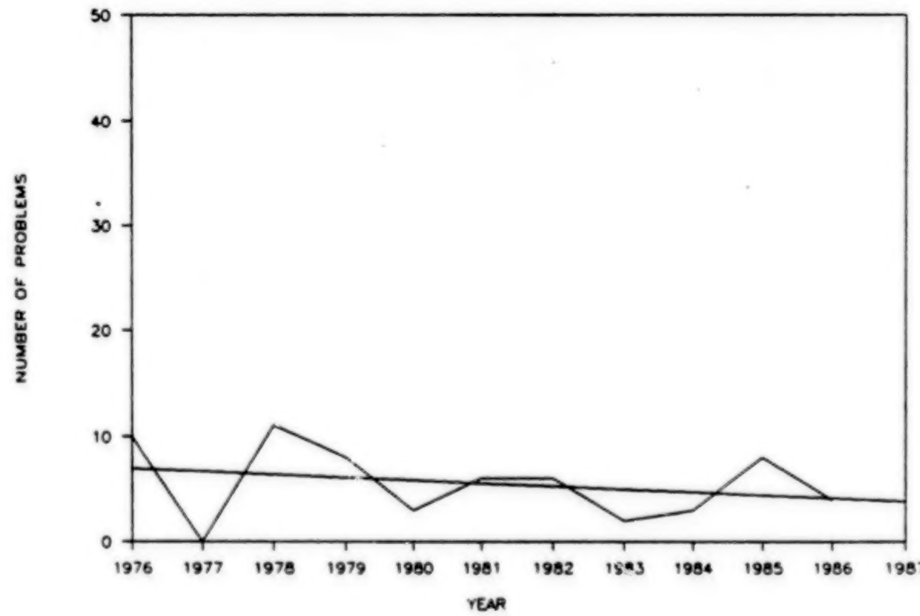


Figure 4.4M. Total MDM Manufacturing Failures

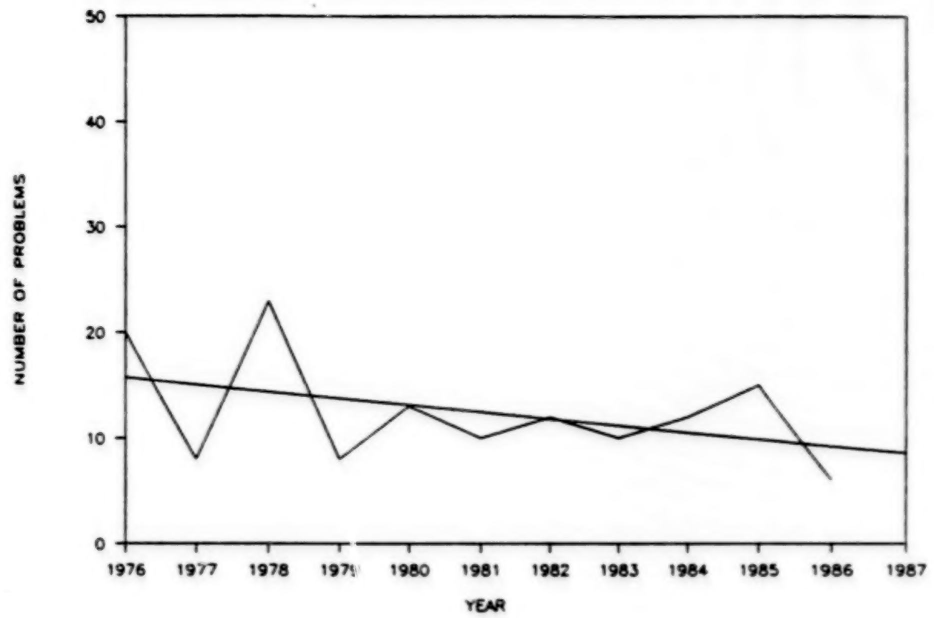


Figure 4.4N. Total MDM Design Failures

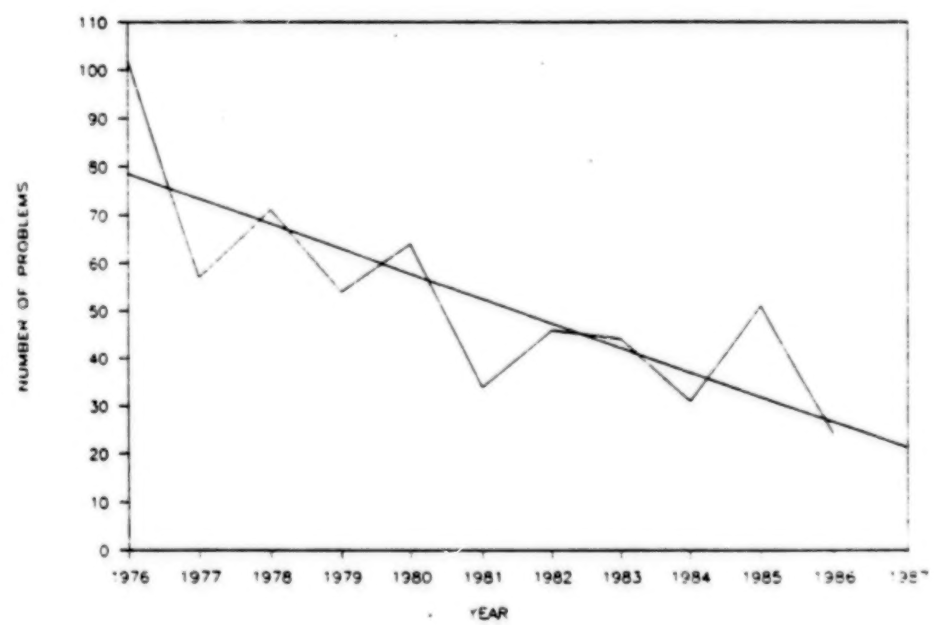


Figure 4.4O. Total MDM EEE Parts Failures

Remarks: General conclusions based on the above trend charts are:

(1) MDM maturity has leveled off at about 50 failures per year, with the majority of failures being EEE. The EEE anomalies have been constant at about 40 per year.

(2) There is no significant trend in manufacturing failures; but the low level of manufacturing failures indicates a maturity in MDM manufacturing.

(3) Design anomalies show a decreasing trend and are low in number.

(4) Despite a large number of failures, only 4 failures have occurred in-flight. Acceptance, certification and field testing have identified nearly all anomalies.

4.4.8.3 Ga As Solar Array Power Supply - Hypothetical Performance Trend. In future long-term manned missions, regenerative H_2/O_2 fuel cell systems may be used. The power producing cells will consume H_2 and O_2; these reactants, in turn, will be created from electrolysis units (power consuming cells) using H_2O, created as a product from the original fuel cell reaction. Solar arrays will drive the water electrolysis units. (Currently, solar cell arrays serve as a power source on the TDRSS and can reach 36 w/kg.)

Hypothetically, assume that a Gallium-Arsenic solar array functions as part of a regenerative power supply system. One kw of continuous power is required from the solar cell. Based on the orbit-sun plane orientation (hence, sunlight exposure), the solar array will produce a peak output of 2 kw, with excess power over 1 kw diverted to storage batteries. Particulate and optical radiation, among other factors, will decrease array lifetime. A 1.6 kw peak output (during an orbit) is deemed a lower bound for adequate power output; and the solar array has a projected lifetime of 2 years, with declining output from 2 to 1.6 kw. The objective is to determine if the declining array output is consistent with a 2-year lifespan (with 1.6 kw the minimum required peak output).

In Figure 4.4P, peak output on a weekly basis was plotted over the initial 40 weeks of the mission. A linear regression (trend line) was fitted to the data and extended beyond the 40-week duration for predictive purposes. Inspection of the trend line shows that the minimal required peak output (1.6 kw) will be reached in less than 104 weeks (approximately 80 weeks).

ARRAY PEAK OUTPUT WATTAGE			
Week	Watts	Week	Watts
1	1998	21	1900
2	1997	22	1910
3	1997	23	1902
4	1996	24	1898
5	1994	25	1892
6	1992	26	1880
7	1994	27	1885
8	1990	28	1880
9	1970	29	1877
10	1977	30	1876
11	1967	31	1870
12	1960	32	1870
13	1950	33	1867
14	1940	34	1850
15	1944	35	1830
16	1940	36	1820
17	1930	37	1809
18	1920	38	1804
19	1926	39	1803
20	1910	40	1800

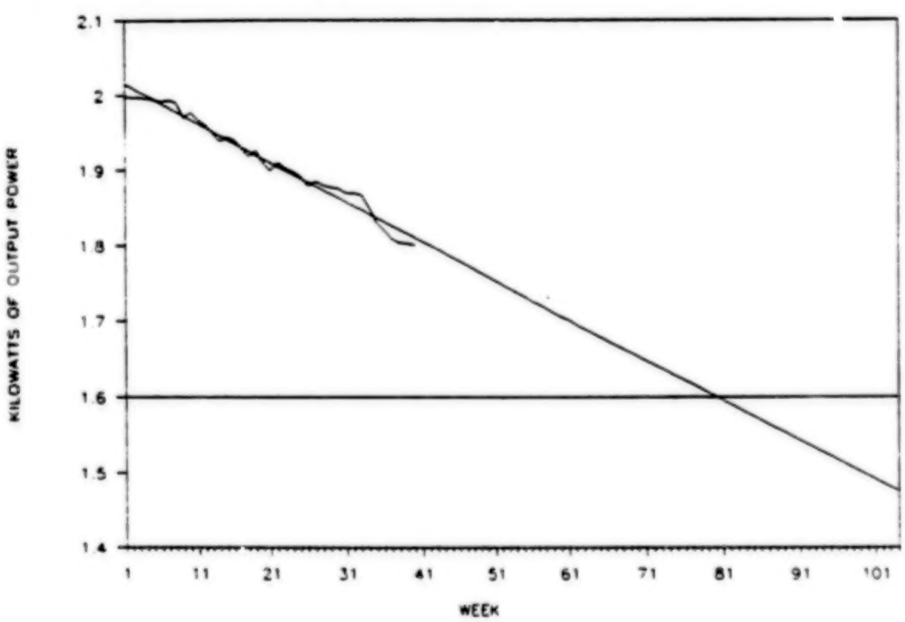


Figure 4.4P. Solar Array Power Output

A more complete analysis is as follows:

Basic Statistics		Equations Used
No. of Observations	= 40	—
Degrees of Freedom	= 38	—
R^2	= .977	(29)
Slope (b)	= -5.24	(17)
Std. Error of Slope	= .13	(38)
Std. Error of Residuals	= 9.61	(37)

Since the R^2 value is very close to 1, there is a significant fit of the trend line to the data. To obtain the value of peak kilowatt output after 104 weeks, one substitutes x = 104 in the equation:

$$\hat{y} = 2020.3 - 5.239x \tag{63}$$

to give

$$\hat{y} = 1475.4 \tag{64}$$

Thus after 2 years of use, the peak output will be below the 1600 watts needed. Solving equation (63) for x (1600 = 2020.3 - 5.239x) gives the predicted time when 1600 watts will be available. Here x = 80 weeks.

The question is: how accurate is the predicted value; i.e., what are the confidence limits on $\hat{y}$ at 80 weeks? The 90% confidence limits for a predicted value $\hat{y}_0$ are given by:

$$(\hat{y}_0) \pm t_{(n-2,\ .95)} \cdot s \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\Sigma\ (x_i-\bar{x})^2} \right]^{1/2} \tag{65}$$

where,

$$s^2 = \frac{\Sigma\ (y_i-\hat{y}_i)^2}{n-2} \tag{66}$$

Here, $t_{(n-2,\ .95)}$ is the t-value corresponding to the 95th percentage point with n-2 degrees of freedom.

The predicted value at 104 weeks is $\hat{y}_{104}$ = 1475.4 watts; and at 80, the predicted value is 1600 watts (the minimum wattage required). We have:

$$\begin{aligned} s &= 9.61 \\ \Sigma\ (x_i-\bar{x})^2 &= 5330 \\ t_{(n-2,\ .95)} &= 1.68 \end{aligned} \tag{67}$$

Substituting these values into expression (65) yields:

$$\hat{y}_{104} \pm 39.7$$
$$\hat{y}_{80} \pm 28.6 \qquad (68)$$

Thus, from expression (68), the 90% confidence interval for the peak output after 104 weeks is:

$$[1435.7,\ 1515.2] \qquad (69)$$

Since this interval does not include 1600, it is not probable that the peak output will be at or above 1600 watts after 2 years.

At 80 weeks (the predicted lifetime using the fitted trend line), from expression (68) the 90% confidence interval is:

$$[1572.6,\ 1629.8] \qquad (70)$$

If that time, x_0, is desired where the lower confidence limit is at or above 1.6 kw with x_0 as large as possible, the confidence interval formula (65) can be used. Substituting $\hat{y}_0 + 2020.3 - 5.239x_0$ for $\hat{y}_0$ in (65), equating to 1600 and solving for x_0 will give this lower limit bounded below by 1600. The calculations give x_0 = 75 weeks. Thus, up to 75 weeks, there is only a 10% likelihood that the solar array will deteriorate below 1600 watts peak output.

The Figures 4.4Q and 4.4R graphically depict these results. Both charts show the data, trend line, and 90% upper and lower confidence limits.

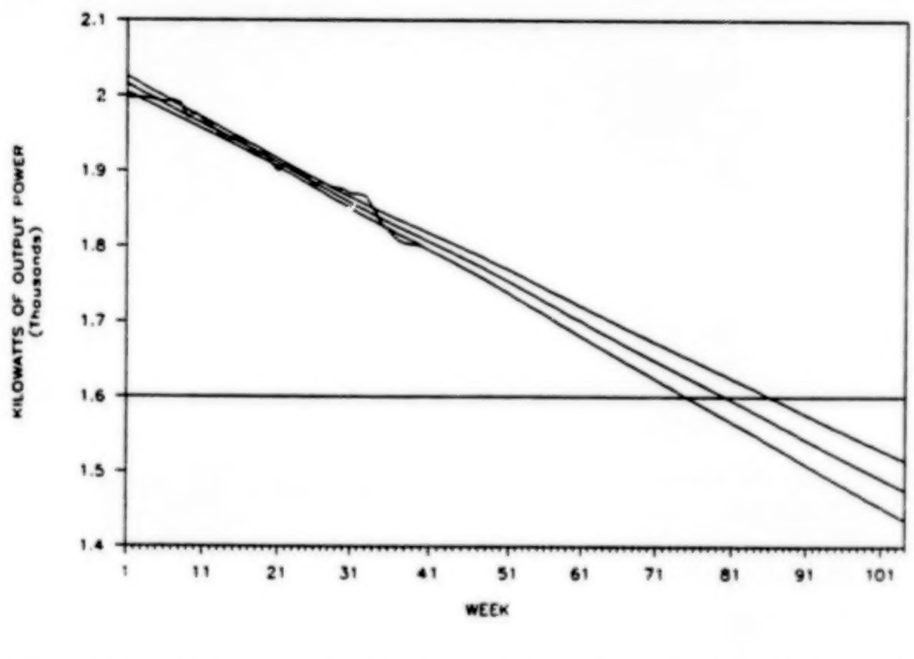


Figure 4.4Q. Solar Array Power Output (with 90% Confidence Intervals)

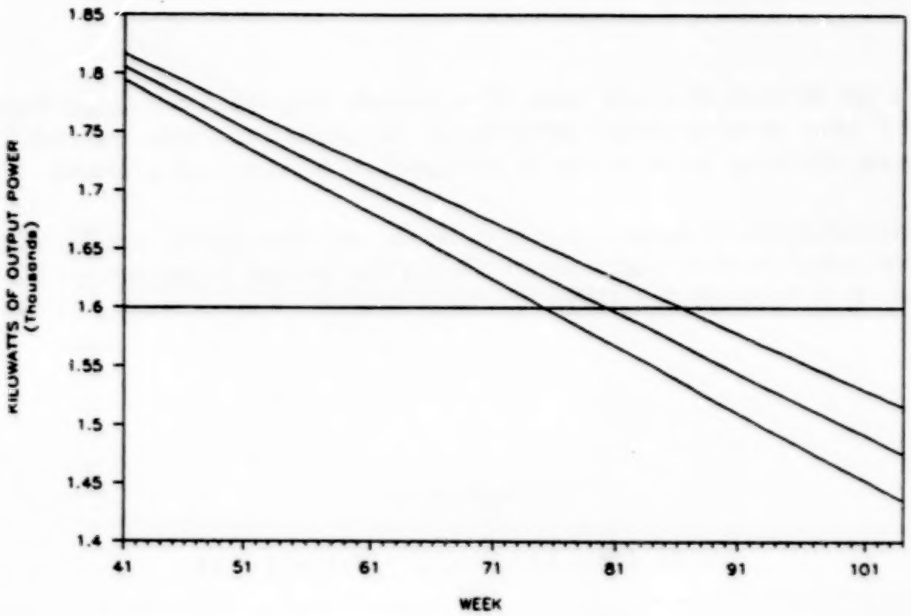


Figure 4.4R. Solar Array Power Output (with 90% Confidence Intervals)

4.4.8.4 High Pressure Turbopump Efficiency: Trend in RPM Levels Using Telemetry Data. The high pressure fuel turbopump (HPFT) operates at approximately 30,000 rpm and boosts the LH_2 pressure from 276 psia (from LPFT discharge) to approximately 6500 psia. Pressure boost is necessary to prevent cavitation in the main combustion chamber.

The following examples use hypothetical telemetry data on HPFT discharge pressure and rpms. The analyses are of two types: one tracks HPFT rpms during a single flight; the other tracks the *maximum HPFT rpms* over a set of past flights. SSME controllers will maintain constant fuel flow pressures for the required thrust levels, adjusting turbine shaft rpms. Trending rpm values for fixed thrust levels will help identify turbopump efficiency declines. (For a declining efficiency, due perhaps to seal leaks, the rpms must be increased above nominal values to maintain required discharge pressures.)

Table 4.4-3 gives HPFT nominal discharge pressure and rpm levels for three SSME-rated power levels:

Table 4.4-3

HPFT PRESSURE AND RPM LEVELS

Power Levels	HPFT	
	Pressure-Discharge	rpm
65%	3953	27,157
100%	6110	34,386
104%	6443	35,361

Telemetry data can be used in several ways. This example uses data on a single flight to check for trends in HPFT rpms at fixed SSME power levels. As noted previously, increasing rpm levels to maintain constant discharge pressure can be indicative of pressure seal problems.

Table 4.4-4 gives observed (hypothetical) rpm levels for the three SSMEs during ascent phase from 1.1 minutes into launch to main engine cutoff. During this period the engines are throttled at 104% of rated power. It is assumed that HPFT discharge pressure was maintained at 6443 psia.

Table 4.4-4

SSME RPM LEVELS (HYPOTHETICAL)

Elapsed Time (minutes)	$SSME_1$ (rpm/10^3)	$SSME_2$ (rpm/10^3)	$SSME_3$ (rpm/10^3)
1.1	35.32	35.37	34.75
1.5	35.33	35.41	35.29
1.9	35.31	35.47	35.56
2.3	35.4	35.51	35.88
2.7	35.36	35.59	35.97
3.1	35.51	36.12	36.40
3.5	35.40	36.08	36.41
3.9	35.49	36.22	36.80
4.3	35.81	36.24	37.10
4.7	35.89	36.37	37.12
5.1	35.70	36.31	37.12
5.5	35.67	36.30	37.20
5.9	35.4	36.19	38.60
6.3	35.82	36.22	38.65
6.7	35.77	36.25	38.70
7.1	35.91	36.17	39.10

By inspection it can be seen that $SSME_1$ and $SSME_2$ rpm levels at 104% power are constant and within the nominal value. It is clear that rpm levels for $SSME_3$ are consistently above the nominal value. Fitting a trend line for $SSME_3$ rpm levels yields the following statistics:

Regression Statistics		Equations Used
R^2 = .956		(29)
a = 34161.84		(17)
b = 671.66		(17)
Standard error (b) = 38.52		(38)
Standard error (y) = 284.11		(37)
Number of Observations	16	—
Degrees of Freedom	14	—

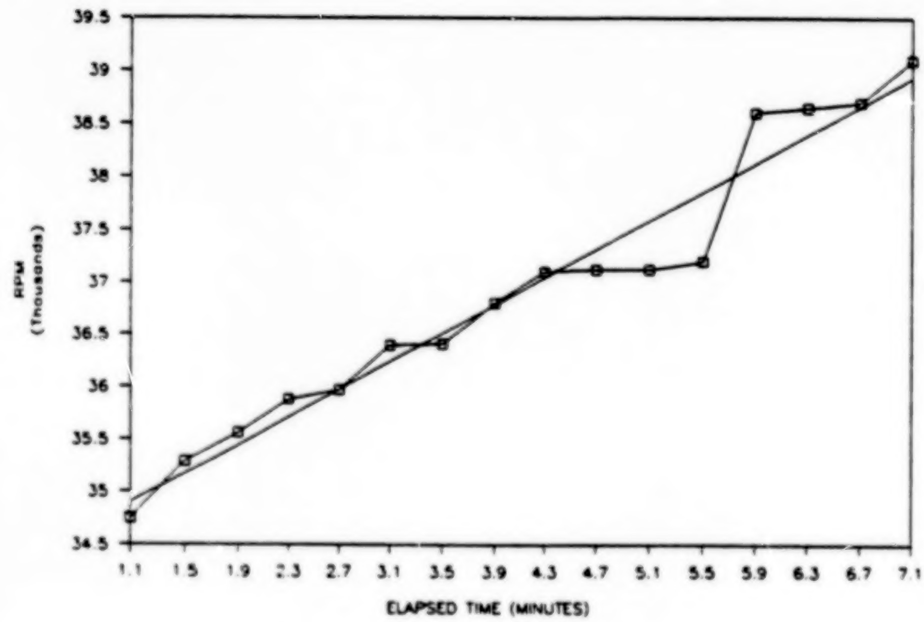


Figure 4.4S. HPFT RPM Levels (at 104% SSME Rated Power)

The trend line slope coefficient is positive (+671.66, hence increasing trend) and the 95% confidence interval for the estimate of β, the true trend line slope, does not include zero. Therefore, there is a statistically significant systematic trend, indicating a problem with maintaining the required turbopump discharge fuel pressure.

4.4.9 Normalization of Trend Data

In the most basic trend applications, we usually have a set of raw data values corresponding to times or time periods. Notationally, these are referenced as the y and x values, respectively. A trend line (or other model) is then fitted to the data. Such a method **assumes** that the magnitude of all observational y-values can be treated equivalently—so that no scaling of y-values by some factor is necessary; however, this is usually not the case. For example, let the x-axis represent three month periods during the years 1986 through 1989 and let the y-axis represent the number of problems occurring (by date of detection) for a given component. A plot of the raw data and subsequent fitting of a trend line implicitly assumes that all time periods are equal with regard to the possibility of problems occurring. However, it may be the case, for example, that during some time periods extensive testing was carried out, while during other periods, little testing of flight experience was accrued. Consequently, during testing periods, the number of problems occurring may be high (in a relative sense), even though there may be no real difference in the actual rate of problem occurrence. Figures 4.4T and 4.4U illustrate the above example.

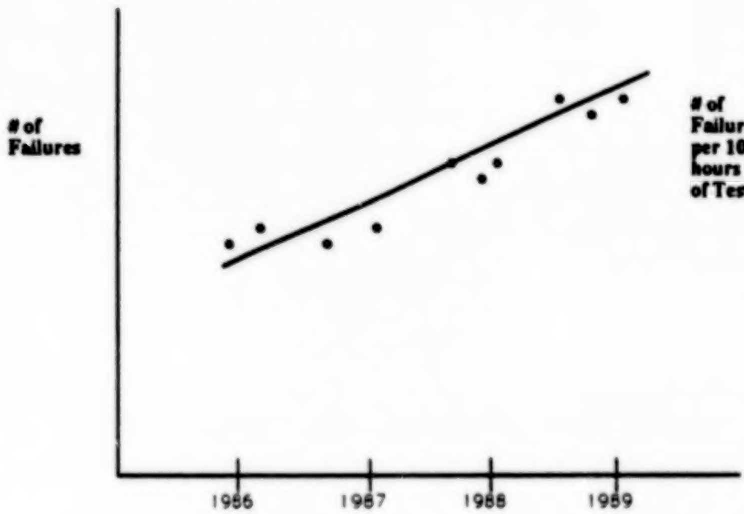


Figure 4.4T. Regression Fit with Raw Data

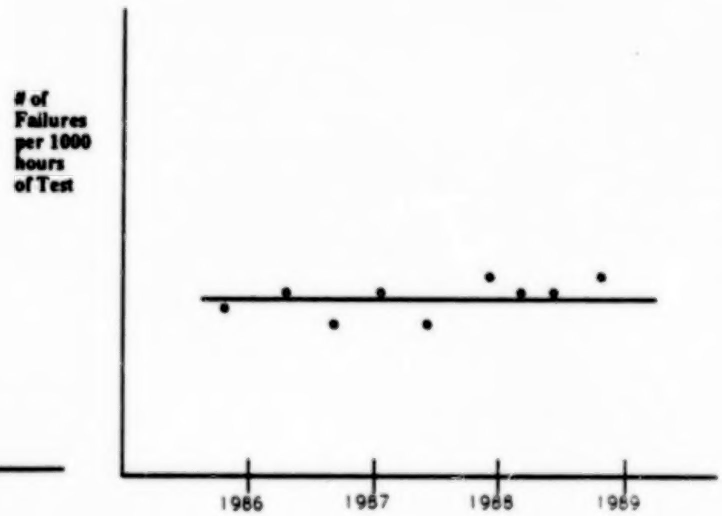
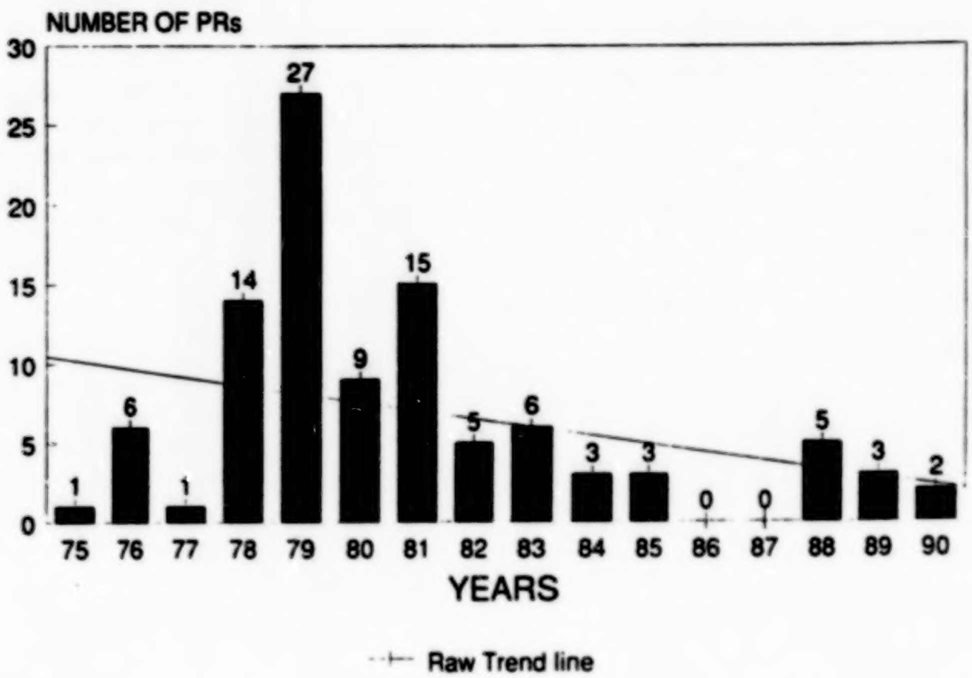


Figure 4.4U. Regression Fit with Data Normalized

By definition, this *scaling* of the observational values is called **normalization**. The choice of a scaling factor or unit is largely a matter of judgement and application. Some typical normalization factors would be: per 1000 direct labor hours (for nonconformance data), per 1000 seconds of test (for SSME component failures), and per vehicle in flow (for supportability issues). The key point is that in trying to determine whether a trend exists, normalization of the data may be necessary to factor out inherent differences (over time) that will unduly influence the magnitude of the observational points. As an example, Figure 4.4V depicts Engine Interface Unit (EIU) Problem Reports (PRs) over a 15-year period. The raw trend line (i.e., line based on nonnormalized data) is plotted.

However, to draw valid conclusions as to whether the actual occurrence of EIU problems is decreasing, approximately constant, or increasing, the amount of test and operational time for the units for each of the years must be known and the number of PRs must be scaled accordingly.



SOURCE: PCASS

Figure 4.4V. EIU Problem Reports Over Time

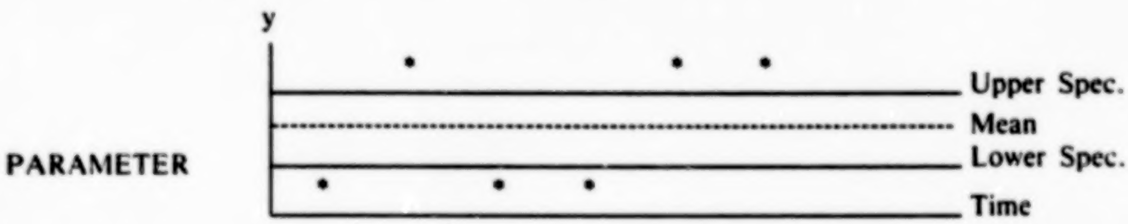
4.5 SMALL SAMPLE METHODS

In some applications, the sample size may be small. Consequently, the uncertainty for any inference estimate will be large (or equivalently, the accuracy or confidence in the estimate is low). In cases where the data is sparse, non-parametric methods are often used.

Non-parametric methods refer to tests that do not depend on assumptions about the underlying distribution from which the sample comes. These tests apply to small samples because no population parameters need be estimated from the sample data. The methods presented are quality control charts and two non-parametric tests: a run-test for systematic trends and a modified Kolmogorov-Smirnov test. These tests can be used to attach a numerical measurement to trend hypotheses; however, in dealing with small data samples, judgement should remain the most important criteria in the data analysis.

4.5.1 Acceptable Level-of-Performance Charts

One of the key ideas in analysis is to compare data to a standard or specification. This standard should be developed during equipment design and verified during prototyping and testing. A simple example might be the diameter tolerance for a rotating shaft; a complex example might be the tolerances on the flow rate of fuel to the SSME. While small data samples often do not provide enough points to conduct a regression analysis, a control chart can quickly reveal trends in their infancy. Also, a regression line may indicate an acceptable mean, yet every point is out of specification, as shown in the following chart:



The mean is well within specification, but the process is not!

The most common method of statistical process control is the use of **control charts**. Although the charts are statistically more effective with large sample sizes or production runs, they can be used for relatively small sample sizes with minimal loss of effectiveness.

Most control charts are developed using $\pm 3\sigma$ control limits, which for a normal distribution includes 99.73% of the population. This results, in an average of 27 cases out of every 10,000 cases giving an erroneous error signal that the process is out of control, when in fact it is not (producer's risk). Although the 3σ standard partly resulted from the limited (by today's standards) computational tools available in the 1920's, it has withstood the test of time and has proven to have been an effective trade-off between *producer's* (α error) and *consumer's* (β error) risk.

Table 4.5-1 is an example of the fraction defective (np) chart. It is applicable to virtually any quality control process. (Refer to figure 4.5A for a graphical representation).

Table 4.5-1

DEFECTS IN SAMPLE LOTS

Sample Lot Number	Sample Size	Number of Defects (np)
1	30	3
2	30	2
3	30	4
4	30	2
5	30	6
6	30	1
7	30	3
8	30	5
9	30	5
10	30	2
	300	33

$n\bar{p} = \frac{33}{10} = 3.3$ (average number of defectives)

$\bar{p} = \frac{33}{300} = .11$ (average fraction defective)

$\sigma = \sqrt{3.3(1-.11)} = \sqrt{2.937} = 1.71$ (standard deviation)

$UCL = n\bar{p} + 3\sigma = 3.3 + 3(1.71) = 8.43$

$LCL = n\bar{p} - 3\sigma = 3.3 - 3(1.71) = -1.83$ (less than zero and therefore is set to zero).

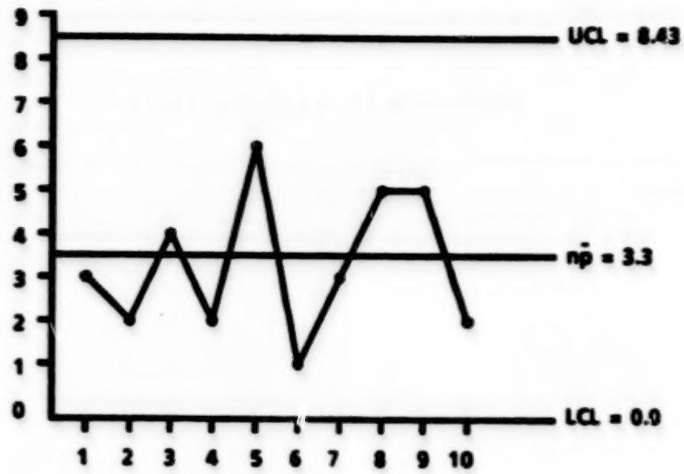


Figure 4.5A. Graphical Representation of the np Chart

Since all of the data points are within the control limits, the process is "in-control." ***It is extremely important to note that in-control does not necessarily indicate within-specification.*** A machining process could be producing a product as indicated by the np chart in the example (in control) where the process specification is $2.5 \pm 1\%$. The product is in control, but not within specification.

In summary, control charts can be designed to fit almost any quality control (or performance trending) application. With the on-line, computer-based power available today, real-time control chart type analyses can be of enormous benefit.

4.5.2 Run Test to See if Observations are Free of Systematic Trends

A Run Test can be used to test whether two samples, $\{x_n\}$ and $\{y_m\}$ come from the same population. The basic idea is to combine or mix the two samples, and then order them by magnitude. On the assumption that both samples come from the same parent population, the ordered sequence of x's and y's should be reasonably mixed. It should not be the case, for example, that most of the x_i appear at the lower end (in order of magnitude) and the y_i appear at the upper end.

By definition, a **run** is a sequence of the same symbol or type. For example, in the sequence x_3 x_1 y_4 y_1 x_2 y_2 y_3 there are 4 runs: a run of 2 x's, a run of 2 y's, a run of 1 x, and a run of 2 y's. When there is a reasonable mix, the number of runs for the $(n+m)$ observations will be large. When the number of runs is small, it is more probable that samples $\{x_n\}$ and $\{y_m\}$ do not come from the same population.

The run test technique can be modified for data obtained sequentially over time to see if the data is *free from systematic trends*. This is the modification that will apply to trending questions. The basic assumption is that all the observations, obtained at different points in time, measure the same

constant quantity, c_0. In performance trending, c_0 is the desired mean performance value. In general, c_0 can be any value, but usually is taken to be the median of the observations $\{x_n\}$. To test whether the values have a constant zero-trend at a specified value, c_0 is taken to be that value.

Now, if all the observations ostensibly measure the quantity c_0, then the observations should hover close to c_0, with equal probability of an observation being above or below c_0. Let A denote an observation above c_0; and let B denote an observation below c_0. Since the observations $\{x_n\}$ are ordered by time, a sequence of As and Bs results. If there is no trend of observations consistently above or below the constant c_0, then the number of runs will be high. However, if the measured values are consistently above (or below) c_0, or consistently below for awhile and then above for awhile, the number of runs will be small. Thus, a small number of runs would indicate that there are systematic trends.

The number of runs, denoted r, has a distribution based on the sample size n and the assumption that the set $\{x_n\}$ consistently measures c_0. This distribution has a mean equal to $\frac{(n+1)}{2}$ and variance

$$\frac{\frac{n}{2}\left(\frac{n}{2}-1\right)}{n-1} \tag{71}$$

The following statistical table gives the critical r-value (denoted $r_{.05}$) at the .05 confidence level. This means that when the number of runs is less than or equal to $r_{.05}$, there is only a .05 probability that the set $\{x_n\}$ consistently measures the constant, c_0. In the trending context, there is only a 5% chance that there is not a systematic trend. Therefore, whenever the computed number of runs, r, is such that $r \leq r_{.05}$, we conclude that there is a systematic trend.

	Critical Values for the Run Statistic* $\alpha = .05$														
n/2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$r_{.05}$	-	-	-	2	3	3	4	5	6	6	7	8	9	10	11

*Comparing against a constant trend value, c_0

Remark: The above run test may be viewed as a quantitative measure to accompany the visual inspection of a control chart.

4.5.3 Kolmogorov-Smirnov Test for Trends

The Kolmogorov-Smirnov test is a goodness-of-fit test used to measure whether a given sample $\{x_n\}$ could come from a certain theoretical model. First the basic ideas in the test will be noted; then modifications that place the test in the context of trend analysis will be given.

A sample of size n, $\{x_n\}$, can be reordered according to magnitude. Assume that the x_i have been so ordered and thus, $x_1 \leq x_2 \leq \ ... \ \leq x_n$. A cumulative distribution function (cdf) $S_n(x)$ is then defined by:

$$S_n(x) = \begin{cases} 0 & \text{if } x < x_1 \\ \dfrac{i}{n} & \text{if } x_i \leq x < x_{i+1} \\ 1 & \text{if } x \geq x_n \end{cases} \tag{72}$$

The function $S_n(x)$ is thus a step function that increases by $\frac{1}{n}$ at each of the observational points.

The idea of the test is to examine the differences between $S_n(x)$, the cumulative distribution for the sample, and any theoretical cumulative distribution function, F(x). If the sample $\{x_n\}$ comes from a population with cumulative function F(x), then one would expect that the difference

$$D_n = \max \{ \mid S_n(x) - F(x) \mid \} \tag{73}$$

over all points is not too large.

Under very general conditions, D_n has a distribution that is independent of any theoretical cumulative distribution. Moreover, the distribution of D_n is sensitive for small samples (n usually $\geq$ 10). This allows testing to determine if the sample could come from a certain distribution or whether that hypothesis can be ruled out.

The Kolmogorov-Smirnov test can be adapted to test for a non-zero trend (or for any specific trend). To illustrate, let $\{y_1, y_2, \ldots, y_n\}$ be the observed values corresponding to times $\{x_1, x_2, \ldots, x_n\}$. The $\{y_i\}$ corresponds to the ordered set $\{x_i\}$ and, in this manner, is already ordered; each y_i is viewed as a frequency at the corresponding time x_i. Therefore, there is a cumulative distribution function, $S_n(x_i)$. S_n is defined as follows:

$$\begin{aligned}
S_0 &= 0 && x < x_1 \\
S_1 &= \frac{y_1}{\Sigma y_i} && x_1 \leq x < x_2 \\
S_2 &= \frac{y_1 + y_2}{\Sigma y_i} && x_2 \leq x < x_3 \\
&\vdots \\
S_{n-1} &= \frac{y_1 + y_2 + \ldots + y_{n-1}}{\Sigma y_i} \\
S_n &= 1 && X_n \leq x
\end{aligned} \tag{74}$$

NOTE: The cdf for the sample of n points is simply the cumulative sum of the frequencies, $\{y_i\}$, normalized by the total frequency, Σy_i.

We want to compare the cdf $S_n(x_i)$ with the theoretical cumulative distribution function determined from a *constant* set of frequencies. The comparison is illustrated in figure 4.5B.

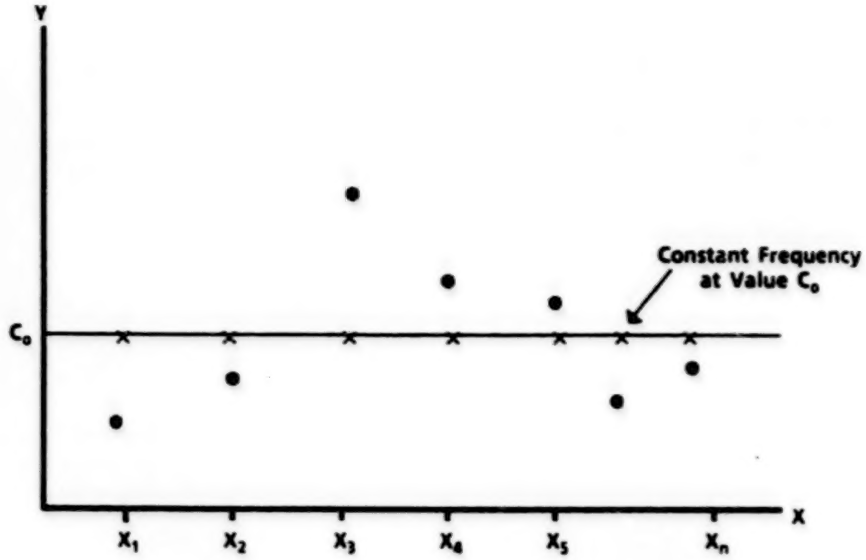


Figure 4.5B. Plot of Observed Values Against a Constant Frequency Distribution

The theoretical cdf F(x) from the uniform distribution of c_0 at each point x_i is representative of a zero-trend, because the values are constant over time. Assuming that the x_i are equally spaced, then F(x) may be defined at the observation points $\{x_i\}$ as:

$$
\begin{aligned}
F(x_1) &= \frac{1}{n} \\
F(x_2) &= \frac{2}{n} \\
&\vdots \\
F(x_i) &= \frac{i}{n} \\
&\vdots \\
F(x_n) &= 1
\end{aligned}
\qquad (75)
$$

Assuming the times are equally spaced for any constant function at the x_i, the cdf for a zero-trend is independent of the specific constant. The value of F(x) at x_i is simply i/n, where n is the total number in the sample.

To use the Kolmogorov-Smirnov test in the trending context, the test statistic is

$$D_n = \max \{ \mid S_n (x) - F(x) \mid \} \tag{76}$$

where S_n and $F(x)$ are as defined previously. The critical values of D_n for various sample sizes, n, appear in Table 4.5-2 at the end of this section. In principle, the differences $|S_n(x) - F(x)|$ are examined for all points x. Since $S_n(x)$ is a step function, only the differences at the *jump* points, x_i, need be computed. Thus, for a sample of size n, 2n differences need to be examined and the maximum taken. The 2n values are $\{|S_n(x_i) - F(x_i)| , |S_n(x_i-1) - F(x_i)|\}$ over the observation points $\{x_i\}$.

For example, suppose the sample size is 5 and a 90% confidence level is adequate. The K-S value, $d_{.05}$, is .5095. Therefore, at the 10% significance level, if the largest deviation between $S_5(x_i)$ and $F(x_i)$ exceeds .5095, then the assumption of a zero trend can be rejected.

As a final point, the table of K-S values shows that for small sample sizes, rather large differences between the cumulative distribution functions must exist for one to conclude that the data must deviate from a constant zero-trend hypothesis.

To summarize the Kolmogorov-Smirnov Test:

Test statistic is $D_n = \max \{ \mid S_n(x) - F(x) \mid \}$,

where

$S_n(x)$ is cdf of observed data and
$F(x)$ is cdf of uniform (constant) distribution.

$$S_n(x) = \frac{y_1 + y_2 + \ldots + y_i}{\Sigma y_i}, \quad F(x) = \frac{i}{n} \tag{77}$$

$$\{\text{for } x_i \le x < x_{i+1}\}$$

A large difference, D_n, indicates that the observed data does not come from a uniform or zero-trend distribution.

Table 4.5-2

PERCENTAGE OF POINTS OF THE KOLMOGOROV-SMIRNOV STATISTIC

The table gives values d_α for specified values of α such that

$$P(D_n \leq d_\alpha) = 1 - \alpha$$

where the Kolmogorov-Smirnov test statistic, D_n, for the sample of size n is the largest deviation between the observed cumulative distribution and the theoretical cumulative distribution.

n \ α	.20	.10	.05	.02	.01
1	.9000	.9500	.9750	.9900	.9950
2	.6838	.7764	.8419	.9000	.9293
3	.5648	.6360	.7076	.7846	.8290
4	.4927	.5652	.6239	.6889	.7342
5	.4470	.5095	.5633	.6272	.6685
6	.4104	.4680	.5193	.5774	.6166
7	.3815	.4361	.4834	.5384	.5758
8	.3583	.4096	.4543	.5065	.5418
9	.3391	.3875	.4300	.4796	.5133
10	.3226	.3687	.4093	.4566	.4889
11	.3083	.3524	.3912	.4367	.4677
12	.2958	.3382	.3754	.4192	.4491
13	.2847	.3255	.3614	.4036	.4325
14	.2748	.3142	.3489	.3897	.4176
15	.2659	.3040	.3376	.3771	.4042
16	.2578	.2947	.3273	.3657	.3920
17	.2504	.2863	.3180	.3553	.3809
18	.2436	.2785	.3094	.3457	.3706
19	.2374	.2714	.3014	.3369	.3612
20	.2316	.2647	.2941	.3287	.3524
21	.2262	.2586	.2872	.3210	.3443
22	.2212	.2528	.2809	.3139	.3367
23	.2165	.2475	.2749	.3073	.3295
24	.2121	.2424	.2693	.3010	.3229
25	.2079	.2377	.2640	.2952	.3166
26	.2040	.2332	.2591	.2896	.3106
27	.2003	.2290	.2544	.2844	.3050
28	.1968	.2250	.2499	.2794	.2997
29	.1935	.2212	.2457	.2747	.2947
30	.1903	.2176	.2417	.2702	.2899
35	.1766	.2019	.2243	.2507	.2690
40	.1655	.1891	.2101	.2349	.2521
45	.1562	.1786	.1984	.2218	.2380
50	.1484	.1696	.1884	.2107	.2260
55	.1416	.1619	.1798	.2011	.2157
60	.1357	.1551	.1723	.1927	.2067
65	.1305	.1491	.1657	.1853	.1988
70	.1259	.1438	.1598	.1786	.1917
75	.1217	.1390	.1544	.1727	.1853
80	.1179	.1347	.1496	.1673	.1795
85	.1144	.1307	.1452	.1624	.1742
90	.1113	.1271	.1412	.1579	.1694
95	.1083	.1238	.1375	.1537	.1649
100	.1056	.1207	.1340	.1499	.1608
≥ 100	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

4.5.4 Examples

4.5.4.1 Example 1: Elevon Servo Actuator - Performance Trend. The elevon servo actuator is used in the shuttle hydraulic system to control orbiter aerosurfaces (elevons). Current applied to the servo linearly increases the secondary delta pressure for hydraulic elevon displacement. The nominal servo breakaway pressure is 50 psia.

Figure 4.5C displays secondary delta pressure as a function of input amperage (in millivolts). The normal operating curve is linear. Post-flight tests on actuator performance shows a decline in secondary delta pressure for a fixed input current level. In general, there is a decreasing trend in performance. (The trend techniques covered in this Standard do not include methods for comparing several curves. A sample of pressures at a *fixed* current should be used to quantify a trend).

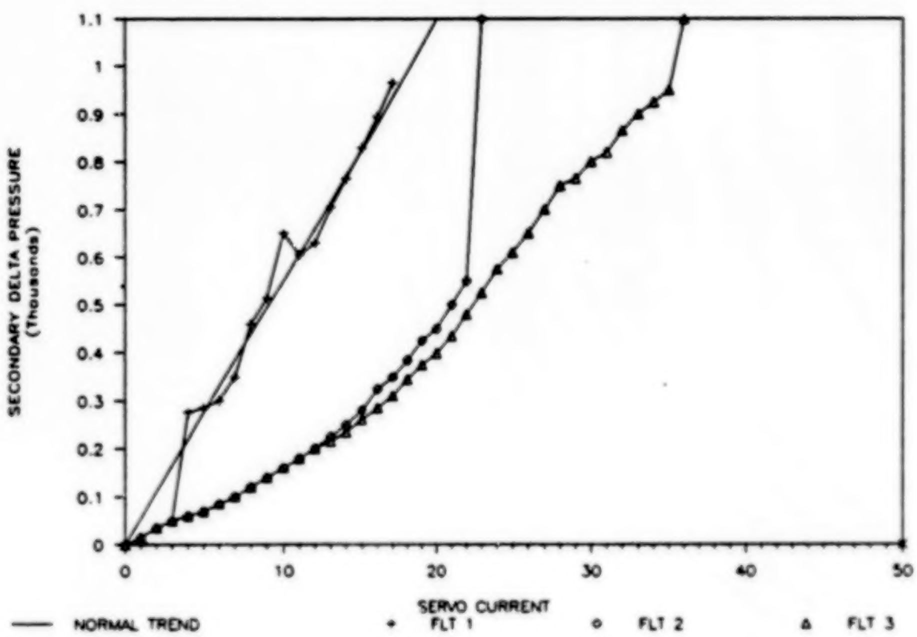


Figure 4.5C. Elevon Servo Actuator (Single Channel Ramp Test)

Figure 4.5D displays simulated data on servo actuator breakaway pressures. Prior to flight, the breakaway pressure in a given actuator is measured. At a breakaway pressure of >450 psi, silting or other problems are occurring; a breakaway pressure of >850 psi requires removal of the unit. In the following control chart, tests 1 through 3 indicate a decreasing level of performance (increasing trend in breakaway pressure). Once the unit reached >450 psi breakaway pressure (test 4), the unit was desilted. Performance is monitored again for subsequent flights. The increasing trend in breakaway pressure is evident. After seven flights, the unit should be replaced.

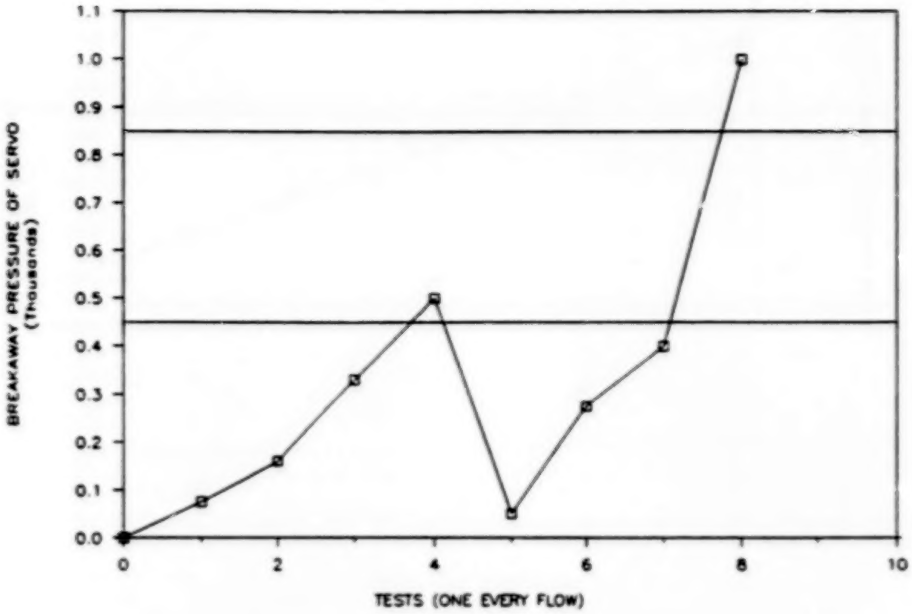
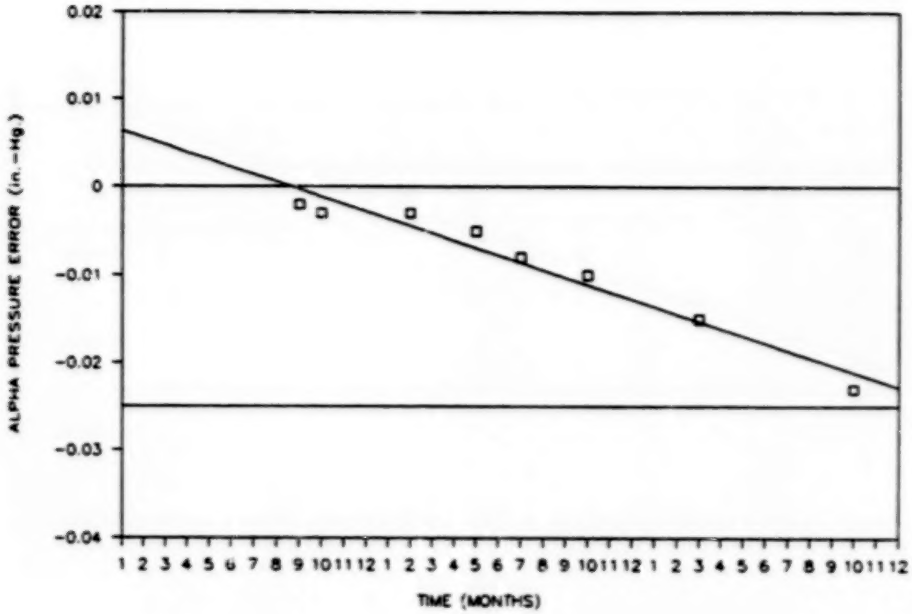


Figure 4.5D. Performance Trend Monitoring (Elevon Servo Actuator)

4.5.4.2 Example 2. Transducer Calibration/Beam Momentum Constancy. The *run* test can be used to test for systematic deviations or trends from a constant value. This example presents actual data for calibration checks on a shuttle air data transducer assembly (ADTA) probe unit. It also presents a hypothetical example on time variation of an ostensibly constant beam momentum used for a high-energy bubble chamber experiment.

a. **ADTA - Calibration Trending.**

The air data transducer assembly is part of the guidance, navigation, and control system. Among other functions, the air probe unit is used in determining the angle and acceleration at reentry. Calibration checks on an ADTA-1 were made throughout the 1982 - 1984 period. Figure 4.5E plots the observed alpha pressures and a fitted trend line to the data.



NOTE: Line at -0.025 indicates the lower limit for *out-of-performance* range.

Figure 4.5E. Air Data Transducer Assembly Calibration Check (ADTA-1)

It is evident that there is a declining trend in calibration from the baseline of 0.0 inches of mercury (in. Hg.). The fitted trend line will give a numerical measure of the decline in calibration (slope = -.00083, standard error of slope = .000076, R^2 = .952). In addition, a run test can be used to give the probability for having the systematic trend of all sampled values below the baseline while still having a calibration distribution that randomly fluctuates above and below the baseline.

The measured 0-baseline calibration values are:

Time		Alpha Pressure (in. Hg.)
Month	Year	
9	1982	-0.002
10	1983	-0.003
2	1984	-0.003
5	1984	-0.005
7	1984	-0.008
10	1984	-0.01
3	1985	-0.015
10	1985	-0.023

Since all observed points lie below the prescribed calibration baseline of 0.0 in. Hg., the series yields a run sequence (with 1 run):

BBBBBBBB

where B denotes *below* the line y = 0.0. Using the run statistic table in section 4.5.3, the critical value at the significance level $\alpha = .05$ and $n/2 = 4$ is $r_{.05} = 2$. (Note that there are eight points, with equal probability of being above or below the constant line; hence using previous notation, for $\{x_n'\}$ and $\{y_m'\}$ as samples, $n' = m' = 4$.) Therefore, the probability of having only one run and no systematic trend is less than .05. Hence, there is a statistically significant declining trend.

However, there is always transducer error, so slight deviations from the 0.0 in. Hg. baseline can be considered errors in measurement rather than true calibration error. Although all measured values are below 0.0, they may fluctuate about the median and cancel out any systematic trends. The run test then can be used to check for systematic trends from the median. This approach also minimizes the error measurement, since the baseline is taken to be the observed median.

The sample median is -.0065. A run sequence about the median is given by:

AAAABBBB

where A denotes *above* the median and B denotes *below*. The sequence has a high degree of order, with only two runs among eight symbols. From the run statistic table, the critical value for $n/2 = 4$ is $r_{.05} = 2$. Hence, the probability of having as few as two runs with no systematic trend is smaller than .05. The conclusion, already seen intuitively, is that the deviations in the ADTA calibration are not due to random errors, but represent a systematic declining trend from the prescribed calibration value.

b. **Beam Momentum Constancy**

The following example illustrates how, in small samples, the *run* test can be used to quantify non-random deviations or trends from a prescribed constant level of functioning. No significant linear trend need exist. Practical applications would include instrument calibrations (as above), sets of tests for leaks prior to flight (i.e., constant pressure to be maintained), and alignment accuracies (e.g., HST fixed head stargazer).

The following data are based on measurements of beam momentum in a bubble chamber experiment. The average momentum of the incident tracks on 20 rolls of film, ordered according to the time of exposure, are:

	GeV/C Units						
[rolls 1-7]	18.61	18.59	18.72	18.69	18.75	19.05	19.06
[rolls 8-14]	19.03	19.04	19.13	19.16	19.19	19.15	19.07
[rolls 15-20]	19.08	19.01	18.91	18.81	18.90	18.89	

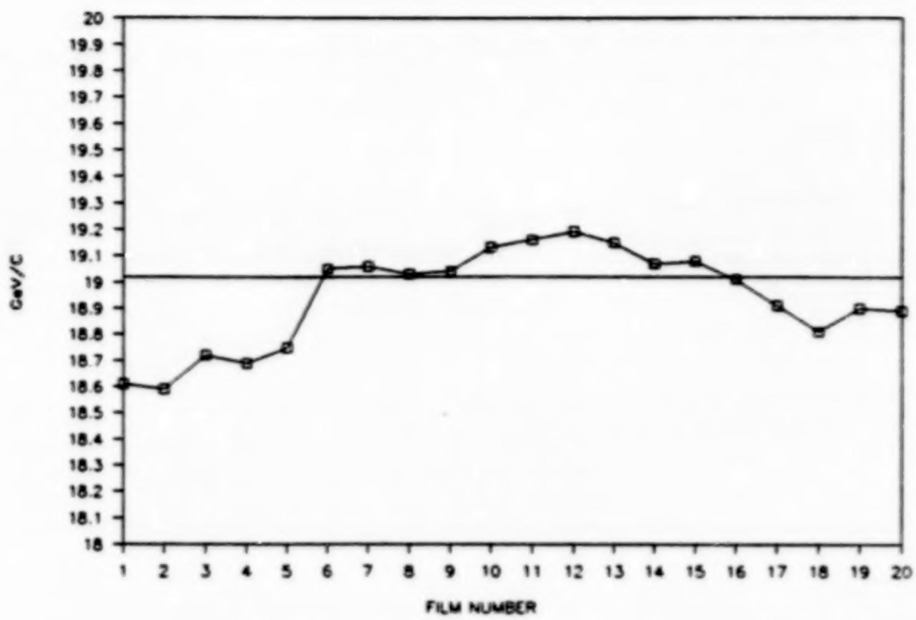


Figure 4.5F. Time Variation of Beam Momentum

Figure 4.5F illustrates that the momentum appears to increase and then decrease during the exposures. A run test is used to determine analytically whether the hypothesis of constant momentum can be rejected. The median value for the observations is 19.02 [(19.03 + 19.01)/2 due to even number of data points]. The classification of values in relation to the mean, where A means above and B means below, yields the series:

BBBBBAAAAAAAAAAABBBBB

This series has 3 runs among the 20 symbols. The table for critical values of the run statistic (here $n/2 = 10$) shows that *less than 6* runs would occur only 5% of the time with a constant momentum beam and measurements deviating randomly about the median. Therefore, the hypothesis that the momentum remains constant is rejected.

4.6 POTENTIAL PROBLEMS IN DETERMINING AND ANALYZING TRENDS

There are a number of subtle and not so subtle errors to avoid in investigating trends. Some of these are noted below.

a. ***The Data is not Normalized.*** In this case, a trend may appear to exist (or not to exist); however, if differences in time periods (that should be *factored out*) are taken into account, there is only an apparent but not a real trend. Figure 4.4T provides such an example.

b. ***Trend Models Become Negative for Future Time Values.*** Data may indicate a declining trend and a trend line (or other model) may have a significant fit; however, at extended time values (i.e., extended x-axis values), the model may take on negative values. In this case, it should not be inferred that the raw values must necessarily decline to zero. Also, in this case, as with any future value prediction, the confidence limits surrounding the predicted value should be computed and reported.

c. ***The Raw Data Contains Numerous Disparate Groupings for Each Time Period.*** By way of example, the observational y-values may be the number of problems associated with a given subsystem. However, these raw numbers represent problems of varying criticality, component, failure mode, or problem. In turn, any conclusions regarding trends at such a top, general level, are probably meaningless. Such a global cut, however, may be valuable in suggesting areas to investigate more closely.

d. ***The Model Has a High Enough R^2 for a Significant Fit—But the Confidence Interval for the Slope Contains Zero.*** With the above conditions, one cannot conclude that the trend is increasing or decreasing. For example, suppose one is trending SSME component contamination, a linear model has a significant fit, and the slope, b, is positive. Assume further that a 90% confidence interval about b is given by $[b - \gamma, b + \gamma]$. One can conclude (prior to actually constructing the interval) that there is a 10% chance that the true slope lay outside of this interval. If the confidence interval does not include zero, one concludes that the trend is increasing. However, if the interval contains zero, there is more than a remote likelihood that the true slope could be flat or decreasing.

Chapter 5

SELECTION OF PARAMETERS (VARIABLES) FOR PERFORMANCE TRENDING

5.1 OVERVIEW/BASIC APPROACH

In performance trending, the values of **variables** (or synonomously, in this section, **parameters**) that either directly or indirectly indicate system/component performance and reliability are examined for trends. Therefore, variables that are *measurable* and related to component performance need to be determined. In analyzing a criticality 1, 1R, or 1S component, it may be the case that no suitable physically measurable variables exist. However, in many cases such parameters and the capability to measure the parameter values over time does exist. The subsequent sections present several methods and criteria for identifying and selecting candidate parameters for performance trending.

The selection of the parameters to be measured and trended is based, in large part, on engineering judgement. There are, however, statistical methods which can be used to select the most highly correlated parameters when suitable raw data on performance is available. A simple correlation coefficient can be computed to see if the parameter and component risk or performance are linearly related. Contingency tables can be used to see if there is *any degree of association* between risk/performance and the potential parameter. And in the case where there are several candidate parameters, and one wants to select those that are the most significant in predicting risk/performance, a stepwise regression technique can be used. Each of these methods is covered in subsequent subsections.

As an overview, determinations for identifying performance parameters should be based on: component criticality; availability of sensor/performance data; trendability of sensor/performance data; and failure/problem history. The FMEA/CIL documents will provide criticality classifications. These documents, together with system assurance and assessment documents, should be examined for failure mechanisms and indicators that can be detected by flight monitoring instrumentation and/or ground checkout and testing. The trendability of data is related to: (1) the establishment of redlines or norms of performance; (2) the accuracy and timeliness of sensor data; and (3) the degree of correlation of the possible parameter with actual component performance.

5.2 SIMPLE CORRELATION

In the case where a performance characteristic or the risk of failure of a component is *linearly related* to another more easily measured variable, a simple correlation test can be used to quantify the degree of linear association. If data is available for correlation, and if a test of association or *correlation* is high, then the more easily measured variable is a reasonable candidate as a performance trending parameter. This section presents the computational aspects of testing for a simple linear correlation.

By definition, the **sample correlation coefficient** r, for a sample of size n,

$$\{(x_1y_1),\ (x_2,y_2),\ \ldots\ ,(x_n,y_n)\} \tag{78}$$

is given by:

$$r = \frac{\sum_{i=1}^{n} (x_i-\bar{x})\ (y_i-\bar{y})}{\left[\sum_{i=1}^{n} (x_i-\bar{x})^2\right]^{1/2} \left[\sum_{i=1}^{n} (y_i-\bar{y})^2\right]^{1/2}} \tag{79}$$

The sample correlation coefficient measures the strength of a linear relationship between the random variables X and Y, even when the data does not come from a bivariate normal distribution. The values of r range between -1 and 1, with r close to 0 implying no correlation and $|r|$ close to 1 implying a high linear correlation.

Since r is based on a *sample*, it is customary to perform a test of significance on the computed sample correlation coefficient. This will give the probability of obtaining the r value (or higher) and having *no* linear relationship between the random variables. In order to perform such a test of significance, the assumption is made that the samples come from a bivariate normal population.

The significance test is based on the fact that the variable Z, defined by

$$Z = \frac{\sqrt{n-3}}{2} \ln\left(\frac{1+r}{1-r}\right) \tag{80}$$

is approximately normally distributed (with mean 0 and variance 1).

A table of the standard normal distribution can then be used to test whether the r value obtained implies that a linear relationship exists.

Suppose, for example, a sample of size 12 is taken, and a correlation coefficient of r = .7 is computed.

Then

$$Z = \frac{\sqrt{12-3}}{2} \ln\left(\frac{1+.7}{1-.7}\right) \tag{81}$$

$$= 2.60$$

Using a table of the standard normal distribution, there is less than a 1% chance of obtaining this z-value and thus this sample correlation coefficient, *while having no correlation*. Hence, the sample shows a high degree of correlation.

5.3 STEPWISE LINEAR REGRESSION

The stepwise linear regression method can be used when there are *several* candidate variables for performance trending and where the aim is to choose one or a subset of these variables. More specifically, the aim is to isolate a subset of variables that will yield an optimal prediction equation with as few variables as possible. In performance trending the dependent variable y usually represents levels of risk of failure (however, y can be any output parameter). The prediction equation is of the form

$$y = a_o + a_1x_1 + a_2x_2 + \ldots + a_nx_n. \tag{82}$$

Many statistical packages (e.g., SAS, SPSS) will perform a stepwise regression and sources in the References section describe the computational method in detail. The key point in the method, however, is as follows:

> The aim in any regression is to find the coefficient estimates $a_o, a_1, \ldots, a_n$ that will minimize the sum of squared differences between the observed value, y, and the predicted value, $\hat{y}$ (over the values of the variables $\{x_i\}$). As noted in section 4.4.3 for the single variable case, the *total variation* or sums of squares can be partitioned into two sums:
>
> $$(y-\bar{y})^2 = \Sigma(y-\hat{y})^2 + \Sigma(\hat{y}-\bar{y})^2 \tag{83}$$
>
> This identity leads to the definition of R^2 as a measure of fit, where R^2 ranges between 0 and 1, and where R^2 measures the amount of total variation that is explained by the regression fit. In a stepwise regression (forward method) the variables are included one-by-one into the predictive equation in the order of highest contribution to explaining the total variance (i.e., those that make R^2 close to 1). In this way, the most significant predictor variables can be isolated.

5.4 TWO-WAY CONTINGENCY TABLES - TESTING FOR INDEPENDENCE BETWEEN TWO CLASSIFICATIONS

A two-way contingency table can be used to measure the degree of association between two variables or classifications. The relevance to parameter selection for performance trending is: (1) the risk of failure can be tested for association with a measurable performance variable; (2) two possible performance parameters can be tested against each other to determine if, in fact, they are independent; if they are highly dependent, only one needs be considered for performance trending; and (3) if a variable is known to be a reliable indicator of risk or performance, but is not easily measurable, another variable may be tested for its degree of association with the original performance variable; a high degree of association (dependence) suggests that the more easily measured variable may be used.

The following discussion presents the basic concepts used in a contingency table analysis. For each observation, one has values for two different variables (e.g. temperature and pressure). The aim is to determine if the values are statistically independent of each other, or whether the value of one variable is associated with the value of the other. A matrix (contingency table) of the values is set up and a chi-square test is used to measure whether any association between variables exists.

In general terms, suppose that two variables or attributes, A and B, are classified into I categories for A, $(A_1, A_2, \ldots, A_I)$ and J categories for B, $(B_1, B_2, \ldots, B_J)$. Variable A categories could be risk

levels (or failure frequencies) for a component and variable B categories could be measured down-line pressure levels.

The counts for observations with combined characteristics A_i and B_j are denoted n_{ij} and can be written as a **contingency table**:

Contingency Table for Two-Way Classification

	B_1	B_2	B_3		B_J	$n_{i\cdot}$
A_1	n_{11}	n_{12}	n_{13}	....	n_{1J}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	n_{23}		n_{2J}	$n_{2\cdot}$
A_3	n_{31}	n_{32}	n_{33}		n_{3J}	$n_{3\cdot}$
	⋮					⋮
A_I	n_{I1}	n_{I2}	n_{I3}		n_{IJ}	$n_{I\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	....	$n_{\cdot J}$	n

The row and column marginal totals are denoted $n_{i\cdot}$ and $n_{\cdot j}$, respectively.

The cell probabilities are estimated by $P_{ij} = n_{ij}/n$. If the cell counts are independently distributed (i.e., there is no association between variables A and B), then the probability for having properties A_i and B_j simultaneously is the product of the probabilities for separate occurrences. Thus, $P_{ij} = P_{i\cdot} \cdot P_{j\cdot}$ and the proportion in row i and column j, n_{ij}/n, should be

$$\frac{n_{ij}}{n} = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} \tag{84}$$

Put another way, if the probability of attribute B_j is independent of any attribute A_i, then the conditional probability $P(B_j|A_i)$ is the same for any A_i, with $P(B_j|A_i) = P(B_j)$, and vice versa. The probability for having both attributes A_i and B_j is

$$P(A_i \cap B_j) = P(A_i) \cdot P(B_j) \tag{85}$$

Assuming the variables A and B are independent, the difference between an observed cell count n_{ij} and an *expected* (based on independence) cell count,

$$n \cdot P_{i\cdot} \cdot P_{\cdot j} = n \left(\frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} \right) \tag{86}$$

should be small. It can be shown that, under the assumption of independence,

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - \frac{n_{i\cdot}\, n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot} \cdot n_{\cdot j}}{n}} \tag{87}$$

is approximately chi-square distributed. The number of degrees of freedom is (I-1)(J-1).

Thus, the more the actual values, n_{ij}, differ from what would be the expected value if the value for row i is independent of being in column j

$$\frac{n_{i\cdot} \cdot n_{\cdot j}}{n} \tag{88}$$

the larger the difference

$$\left(n_{ij} - \frac{n_{i\cdot}\, n_{\cdot j}}{n}\right)^2 \tag{89}$$

The summation over all such differences would then be large. Consequently, a larger chi-square value indicates a high degree of association. Conversely, a chi-square value close to zero indicates that the variables are independent.

The example in the next section demonstrates the use for and computations in a contingency table/chi-square test of association.

5.5 EXAMPLE: LUBRICATION FLUID PARTICULATE CONCENTRATION AS AN INDICATION OF TURBINE SHAFT WEAR

Turbine shaft play in a shuttle component is thought to be associated with the metal particulate concentration of the component's lubrication fluid. In general, analysis of used lubricant can be used to detect unusual wear and to predict impending failures. A spectrometer (based on emission/absorption principles) is used to determine the amounts of wear-metals or elements in a lubrication fluid sample.

To ascertain whether lubrication particulate concentration may be used as a performance indication of shaft wear, data on shaft wear and particulate concentration were obtained. The following contingency table summarizes this data.

Shaft Wear	Particulate Concentration (mg/ml)					
	$10^{-4} < x$	$10^{-4} \leq x < 10^{-3}$	$10^{-3} \leq x < 10^{-2}$	$10^{-2} \leq x < 10^{-1}$	$x \geq 10^{-1}$	
Excessive Shaft Wear	1	3	7	4	10	25
Minor Shaft Wear	5	6	7	14	6	38
No Shaft Wear	15	11	9	10	4	49
	21	20	23	28	20	112

The "expected" cell counts, assuming there is no association, are:

4.69	4.46	5.13	6.25	4.46
7.12	6.79	7.8	9.5	6.79
9.19	8.75	10.06	12.25	8.75

Carrying out the computation for the measurement of the difference between observed and expected values by equation (87), the chi-square value is:

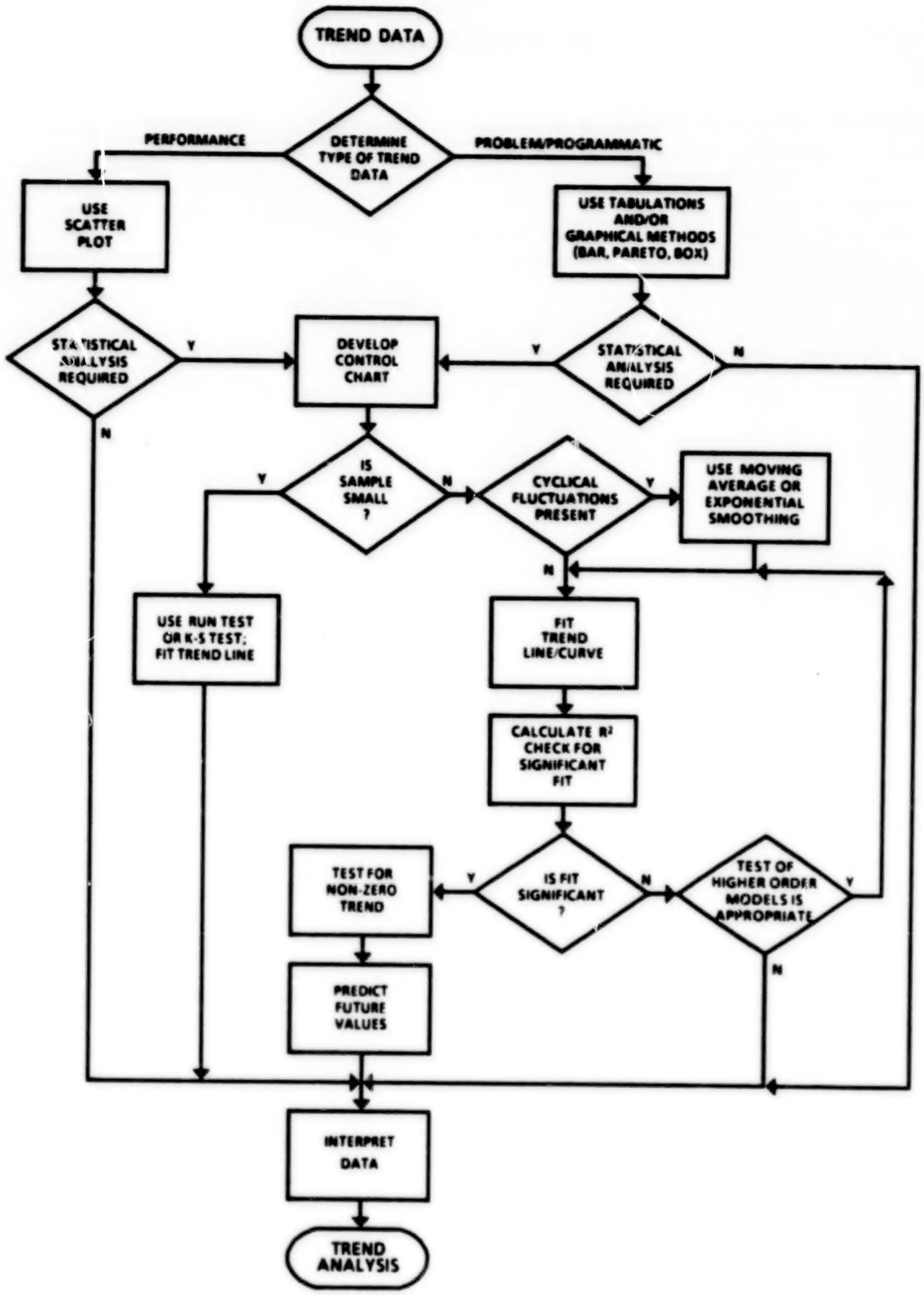
$$\chi^2 = 22.12 \tag{90}$$

with $(3-1)(5-1) = 8$ degrees of freedom. Using a chi-square table, the probability for no association is less than .5%. Therefore, the metal particulate concentration of lubricant fluid would be a strong indication of performance/reliability. The particulate concentration parameter would be a good candidate for performance trending of the component's turbine shaft wear.

Chapter 6

TECHNIQUE SELECTION

The graphical, descriptive statistics and analytical techniques described in this document should be employed to identify and evaluate potentially hazardous or otherwise significant trends. In many cases, simple descriptive statistics and graphical representations of the data will suffice in trend identification/evaluation. When a numerical measurement of trend is needed, the analytical regression and small sample techniques apply. The accompanying flowchart depicts the general progression in selection of techniques for a trend analysis.


TREND DATA
DETERMINE TYPE OF TREND DATA
PERFORMANCE
PROBLEM/PROGRAMMATIC
USE SCATTER PLOT
USE TABULATIONS AND/OR GRAPHICAL METHODS (BAR, PARETO, BOX)
STATISTICAL ANALYSIS REQUIRED
Y
N
DEVELOP CONTROL CHART
STATISTICAL ANALYSIS REQUIRED
IS SAMPLE SMALL ?
CYCLICAL FLUCTUATIONS PRESENT
USE MOVING AVERAGE OR EXPONENTIAL SMOOTHING
USE RUN TEST OR K-S TEST; FIT TREND LINE
FIT TREND LINE/CURVE
CALCULATE R² CHECK FOR SIGNIFICANT FIT
TEST FOR NON-ZERO TREND
IS FIT SIGNIFICANT ?
TEST OF HIGHER ORDER MODELS IS APPROPRIATE
PREDICT FUTURE VALUES
INTERPRET DATA
TREND ANALYSIS

REFERENCES

1. Bury, K., *Statistical Models in Applied Science;* John Wiley, New York; dated 1975.

2. Chou, Y.; *Statistical Analysis;* Second Ed., Holt, Rinehart and Winston; dated 1975.

3. Draper, N.R. and Smith, H., *Applied Regression Analysis;* Second Ed., John Wiley, New York; dated 1981.

4. Halpern, S., *The Assurance Sciences - An Introduction to Quality Control and Reliability;* Prentice-Hall, New Jersey; dated 1978.

5. Snedecor, G.W., and Cochran, W.G., *Statistical Methods;* Sixth Ed., The Iowa State University Press, Ames, Iowa; dated 1978.

6. Wadsworth, H. Jr., Stephens, K.S., and Godfrey, A., *Modern Methods for Quality Control and Improvement;* John Wiley, New York; dated 1986.

7. Natrella, M.G., *Experimental Statistics, National Bureau of Standards Handbook 91,* U.S. Government Printing Office, Washington, D.C.; dated 1963.

TREND ANALYSIS TECHNIQUES CHANGE SUGGESTION FORM

From:______________________ Center/Code/Contractor:____________________

Phone:_____________________ Date:___________________________________

Revision: A Chapter:_______ Section:_______ Page:________

Suggested Change:

Rationale:

Forward to Peter J. Rutledge
Manager, Trend Analysis Policy
NASA Headquarters, Code QT
Washington, DC 20546

END

DATE

FILMED

JUN 25 1991